Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy

Dissertation

Submitted to Northcentral University

Graduate Faculty of the School of Education
in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

JENNY GRANT RANKIN

Prescott Valley, Arizona

September, 2013

APPROVAL PAGE

Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy

by

Jenny Grant Rankin

Approved by:

Howard L. Jacobs Ph.D. 9/19/13

Chair: Dr. Howard Jacobs, Ph.D. Date

Member: Dr. Paul Burd, Ph.D.

Member: Dr. Scherrine Davenport, Ph.D.

Certified by:

Dr. Cindy K Guillaume 9/19/2013

School of Education Dean: Dr. Cindy Guillaume, Ed.D. Date

Abstract

There is extensive research on the benefits of making data-informed decisions, but research also contains evidence many educators incorrectly interpret student data. Meanwhile, the types of detailed labeling on over-the-counter medication have been shown to improve use of *non*-medication products, as well. However, data systems most educators use to analyze student data usually display data without supporting guidance concerning the data's proper analysis. In this dissertation, the data-equivalent to over-the-counter medicine is termed *over-the-counter data*: essentially, enlisting medical label conventions to pair data reports with straightforward verbiage on the proper interpretation of report contents. The researcher in this experimental, quantitative study explored the inclusion of such supports in data systems and their reports. The cross-sectional sampling of 211 educators of varied backgrounds and roles at nine elementary and secondary schools throughout California answered survey questions regarding student data reports with varied forms of analysis guidance. Respondents' data analyses were found to be 307% more accurate when a report footer was present, 205% more accurate when an abstract was present, and 273% more accurate when an interpretation guide was present. These findings and others were significant and fill a void in field literature by containing evidence that can be used to identify how data systems can increase data analysis accuracy by offering analysis support through labeling and supplemental documentation. Recommendations for future research include measuring the impact over-the-counter data has on data analysis accuracy when all supports are offered to educators in concert.

Keywords: abstract, analysis, data, data-driven decision-making, DDDM, data-informed decision-making, data system, data warehouse, footer, ICT, interpretation guide, report

Dedication

This dissertation is dedicated to the loving memory of my father (Donald A. Grant), who was the first great teacher in my life and my hero, and to my mother (Nancy S. Grant), who always models the altruism, intellect, and humor that characterize the best educators.

Acknowledgements

I owe deep gratitude to my mentor and committee chair, Dr. Howard Jacobs, Ph.D. Without his expertise, guidance, and unwavering encouragement, this dissertation would not have met the caliber I had hoped for it. I am particularly grateful for his humor, which always brought much-needed levity and understanding to the otherwise-arduous doctoral process. I am also grateful for the valuable feedback from my committee members.

I thank those such as Dr. Jeffrey C. Wayman who continue to talk about the role data systems and reports play in the effectiveness of educators' data use. Their offers of time were appreciated, as are the contributions they make to a field of literature vital to the improvement of educational technology.

Special thanks go to Dr. Linda Orozco, who encouraged me to pursue my Ph.D., and to those who procured study participants. This includes EdSurge, which so generously shared the participation opportunity in its wonderful newsletters. This also includes educators who graciously encouraged others to invite me onto their campuses.

Finally, I extend sincere thanks and respect to the educators who participated in this study, as well as to those who so generously arranged for it to take place at their own school sites. Having spent most of my career as an educator, I understand how precious time is for those who work tirelessly on behalf of students, so I am especially grateful for their gifts of time. My greatest hope for this dissertation is that its results will be used to improve the manner in which data systems communicate data to educators so as to better assist them in helping students. Thus, participating in this study was yet one more way these participants gave selflessly for kids. Our future is a bright vision when we have such champions for students in our schools.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

In cases where someone is not receiving medicine directly from a doctor, the information on over-the-counter medication's label is crucial to consumer safety (DeWalt, 2010). The medicine's purpose, ingredients, dosage instructions, and dangers are all outlined on a detailed label (Kuehn, 2009). With such guidance, patients may take over-the-counter medication with the goal of improving wellbeing while a doctor is not present to explain how to use the medication.

Label conventions can result in improved understanding on non-medication products, as well (Hampton, 2007; Qin et al., 2011). Thus, in the way over-the-counter medicine's proper use is communicated with a thorough label and sometimes with added documentation, a data system used to analyze student performance can include components to help users better comprehend the data it contains. A data system, also referred to in education as a student data system, is software that provides student data to educators in a digestible, report-based format (Wayman, 2005). Educators use data systems to make decisions that impact students (VanWinkle, Vezzu, & Zapata-Rivera, 2011). No or poor medication labels have resulted in many errors and tragedy, as people are left with no way to know how to use the contents wisely (Brown-Brumfield & DeLeon, 2010). Yet many data systems display data for educators without sufficient support to use their contents – data – wisely (Coburn, Honig, & Stein, 2009; Data Quality Campaign [DQC], 2009, 2011; Goodman & Hambleton, 2004; National Forum on Education Statistics [NFES], 2011). Feedback is considered one of the most powerful influences on student learning and achievement, but this impact can be negative if the performance feedback is not provided in the best way (Hattie & Timperley, 2013).

This paper features an exploration of the concept of over-the-counter data: essentially, the prospect of improving educators' data use by embedding data usage guidance within the data systems they are using to analyze data, just as over-the-counter medication is packaged with usage guidelines. Background is provided as to why the research topic is timely and of interest. The problem statement contains evidence of educators' high error rate when drawing data-based conclusions, their tendency to analyze data while alone and without potential supports outside of the data system, and the lack of analysis support currently within most data systems. The purpose statement and research questions reflect the quantitative study goal of investigating the degree to which such usage guidance can help. Finally, the nature and significance of the study are explained, followed by definitions of key terms and a summary of the study.

Background

The Food and Drug Administration (FDA) requires the pharmaceutical industry to accompany over-the-counter medication with textual guidance regarding its use and to also provide solid evidence on how effective its labeling is in reducing errors, deeming it negligent to do otherwise (DeWalt, 2010). Data systems are commonly used to generate data reports, yet research on aspects of report format and system support that could enhance analysis accuracy is scarce (Goodman & Hambleton, 2004). Research that was devoted to data system and report format limits this exploration to participants' preferences and participants' perceived value of supports. However, user preference can be the opposite of the reporting format that actually renders the more accurate interpretation (Hattie, 2010).

This study was used to examine how effective varied analysis supports are in improving data analysis accuracy, and it did not rely on participants' preferences or perceived value of supports. It was thus unique in determining the specific extent to which each form of analysis guidance improves analysis accuracy, and rendering examples and templates for real-world implementation. The findings of this study filled a gap in education field literature by containing evidence that can be used to identify whether, how, and to what extent data systems can help increase educators' data analysis accuracy by providing analysis support within data systems and their reports. Improvements data system and report providers make in light of this study have the potential to improve the accuracy with which educators analyze the data generated by their data systems. More accurate data analyses will likely result in more accurate data-informed decision-making for the benefit of students.

Statement of the Problem

The problem investigated was educators make data analysis errors impacting students, yet data systems and reports do not include analysis help, and it was undecided whether adding supports to data systems can reduce the number of analysis errors. Data-informed decisions can improve learning (Sabbah, 2011; Underwood, Zapata-Rivera, & VanWinkle, 2010; Wohlstetter, Datnow, & Park, 2008). Educators worldwide test students, distribute score reports, and expect stakeholders to make improvements based on these reports (Hattie & Brown, 2008). Most educators have access to data systems to generate and analyze score reports (Aarons, 2009; Herbert, 2011).

Unfortunately, educators do not use this data correctly, and there is clear evidence many users of data system reports have trouble understanding the data (Hattie, 2010;

National Research Council, 2001; Wayman, Snodgrass Rangel, Jimerson, & Cho, 2010; Zwick et al., 2008). For example, in a national study of districts known for *strong* data use, teachers incorrectly interpreted 52% of data (U.S. Department of Education Office of Planning, Evaluation and Policy Development [USDEOPEPD], 2009). Few teacher preparation programs cover topics like assessment data literacy (Halpin & Cauthen, 2011; Stiggins, 2002), most people analyzing data received *no* training to do so (DQC, 2009; Few, 2008), and human biases compromise judgment and complicate decision-making processes (Kahneman, 2011).

Data use impacts students, and misunderstandings when using data systems can cripple data use in school districts (Wayman, Cho, & Shaw, 2009). Yet labeling and tools within data systems to assist analysis are uncommon, even though most educators analyze data alone (USDEOPEPD, 2009). There is a clear need for research identifying how reports can better facilitate correct interpretations by its users (Goodman & Hambleton, 2004; Hattie, 2010). The power of data systems that generate these reports will not be realized until researchers contribute to improving data system design to improve analysis (DQC, 2011).

Purpose of the Study

The purpose of this experimental quantitative study, conducted in a laboratory environment, was to facilitate causal inferences concerning the degree to which including different forms of data usage guidance within a data system reporting environment can improve educators' understanding of the data contents, much like including different forms of usage guidance with over-the-counter medication is needed to properly communicate how to use its contents. Independent variables included brief, cautionary

verbiage in report footers, report-specific abstracts, and report-specific interpretation guides. The dependent variable was accuracy of data analysis-based responses. The researcher explored three data analysis supports provided by a data system, each framed in two different formats, by presenting 211 elementary and secondary educators in ethnically and culturally diverse southern California with different versions of the same two student achievement data report environments. Each of these report sets fit into one of the following treatment categories (a) no added analysis support; (b) analysis support by way of footers directly on the reports, which were offered in two different framing styles; (c) analysis support by way of abstracts, which accompanied the reports and were offered in two different framing styles; and (d) by way of interpretation guides, which accompanied the reports and were offered in two different framing styles (see *Appendix C* for reports and handouts). The researcher then compared the results of educators using data system reports embedded with data analysis guidance in the varied formats noted above (a-c). Participant responses were collected through a web-based questionnaire crafted and administered in Google Docs, taking advantage of the Google Form feature, and involved groups of no more than 30 respondents at each administration time at each participant's school site. Data was collected at one point in time for each participant within a one-month research window. Findings from this research are suited to identify whether data systems used by educators can help prevent common analysis mistakes by providing analysis support within the interface and the reports they are used to generate.

Theoretical Framework

This research study fell within the conceptual and theoretical area of data-informed decision-making as a means of raising student achievement, as it included an

exploration of how data systems can improve educator accuracy when performing the data analysis step of data-informed decision-making. Data use can lead to insight into students' abilities and to decisions to improve instruction (Underwood et al., 2010). Research review indicates using data to inform instructional decisions can result in greater student achievement (Lewis, Madison-Harris, Muoneke, Times, 2010; Wayman, 2005; Wohlstetter et al., 2008). Thus educators realize data can be the foundation for action toward school improvement (Sabbah, 2011; Supovtiz & Klein, 2003).

Largely due to the No Child Left Behind (NCLB) Act of 2001 that increased pressure on educators to raise student achievement, data interpretation has become increasingly vital to school reform (Minnici & Hill, 2007). Worldwide, nations and U.S. states use some form of national or state-wide testing; distribute score reports to students, parents, educators, and/or government; and expect stakeholders to learn from these reports and use them for data-informed decision-making (Hattie & Brown, 2008). However, even the name of the premise these stakeholders are employing – data-*informed* decision-making – indicates it relies on the understanding that the data is being used to *inform* decisions, not *mis*inform them. Misunderstandings about how to use data and a data system can cripple data use in a school district and cause low data system use rates and resistance to data (Wayman et al., 2009).

Frequent problem. The value of data-informed decision-making is negated when educators do not analyze the data correctly when using is to make decisions. Data is useless if we cannot understand it (Few, 2008). Unfortunately, not all educators have the skills needed to successfully use data to inform decisions, and having data does not mean it will be used properly (Marsh, Pane, & Hamilton, 2006). Few educators automatically

know how to use available data effectively (DQC, 2009), and many educators experience difficulties just trying to understand what it means (Goodman & Hambleton, 2004; Hambleton, 2002; Hattie, 2010; NRC, 2001). Educators must be skilled at using data daily to improve student learning, yet many are not (Zwick et al., 2008).

Teachers have frequent difficulties using data, express a need for easier ways to use data, and are overwhelmed by data, (Wayman et al., 2010). For example, teachers have difficulty using data systems due to varying technological sophistication levels when it comes to using the data system to interpret student data, even amongst teachers who serve as assessment coaches to their peers (Underwood, Zapata-Rivera, & VanWinkle, 2008). The problem is not restricted to teachers. Stakeholders at all levels have trouble interpreting data, such as principals who are intimidated by data and need training, and teacher coaches who are not tech-savvy and have trouble sharing assessments and data system knowledge with teachers (Underwood et al., 2008). State-level stakeholders are also at varying stages of being able to actually analyze the data that data systems display (Minnici & Hill, 2007). Even at the state level, stakeholders are not using student data effectively (Halpin & Cauthen, 2011). Yet if data system users do not understand how to properly analyze data, the data will be used incorrectly if it is used at all (NFES, 2011).

Contributing factors. Multiple variables can lead to flawed data-informed decision-making. For example, educators' incomplete understanding of statistics can lead them to draw false conclusions from data (Marsh et al., 2006). Many teachers and administrators do not know fundamental analysis concepts, and 70% have never taken a college or post graduate course in educational measurement (Zwick et al., 2008). Few

teacher preparation programs cover topics like state data literacy (Halpin & Cauthen, 2011; Stiggins, 2002). Training programs for teachers have generally not addressed data skills and data-informed decision-making (USDEOPEPD, 2011). In fact, most people responsible for analyzing data have received no training to do so (DQC, 2009; Few, 2008).

Two solution theories: professional development and staff. Most educators are eager to analyze and then act on the data they see, but interpretations require knowledge and understanding (Hattie, 2010; van der Meij, 2008). Two theories dominate most literature concerning how best to equip educators with the knowledge and understanding needed to correctly interpret and use data. One of these theories is that professional development (PD) can improve educators' data analysis accuracy (Lukin, Bandalos, Eckhout, & Mickelson, 2004; Sanchez, Kline, & Laird, 2009; Zwick et al., 2008). The other prevailing theory is staff resources such as site leaders, data teams, data experts, and/or instructional coaches can improve educators' data analysis accuracy (Bennett & Gitomer, 2009; McLaughlin & Talbert, 2006). While there is research-based merit to both these theories (see *Chapter 2: Literature: Controversy Concerning the Best Way to Improve Data Analysis Accuracy* for specifics), there are also limitations to both approaches (see *Chapter 2: Literature: Supports Outside of Data Systems Are Not Enough* for specifics). Even when educators benefit from employing these two solutions, students deserve for educators to use *all* possible supports for improved analysis accuracy in an effort to completely eliminate – rather than merely reduce – their analysis errors when using those data analyses to make decisions.

Third solution theory: analysis tool improvement. The avenue for analysis accuracy this study was used to explore concerns the data system reports that display the data educators are interpreting. The role the data system plays in analysis accuracy is largely ignored by research literature, but not entirely. There is clear evidence many users of data reports have trouble understanding and interpreting data as it is displayed in these reports (Goodman & Hambleton, 2004; Hambleton, 2002; Hattie, 2010). For example, teachers do not understand or value some data when viewing it in data system reports (Underwood et al., 2008). Teachers need additional help understanding measurement concepts and statistical terms, and adding information to reports can provide this help (Zapata-Rivera & VanWinkle, 2010). Problems analyzing data in data systems and their reports extend to other educational roles, as well. Although administrators are increasingly asked to make data-informed decisions, they have trouble understanding data presented in reports (VanWinkle et al., 2011). Administrators misunderstand the meanings of symbols and terms used in assessment reports and are often confused by the reports' complexity, and the reports district administrators are charged with using are presented in ways that are hard for them to read and interpret (Underwood et al., 2010). Reports are rarely available in formats district administrators can use (Coburn et al., 2009; Underwood et al., 2010). Even stakeholders such as state politicians, superintendents, and education reporters frequently misunderstand and misinterpret national assessment score reports (Hambleton & Slater, 1996; VanWinkle et al., 2011).

The U.S.'s NCLB Act of 2001 led to reports for multiple subjects being distributed at the state, district, school, subgroup, and student levels for parents and teachers of 22 million students per year, yet the reports are not in accordance with any

nationally recognized reporting standards, whereas not all users of these reports are as test sophisticated as they need to be to use them (Hattie, 2010). Quality control for reporting of student data relates largely to mistakes relating to score validity, such as test examiner or computer errors, and little to do with report design and analysis errors (Allalouf, 2009). This is unfortunate, as the manner in which data is presented significantly impacts the decisions that data is used to make (Thaler & Sunstein, 2008). Researchers reveal many educators have difficulty understanding the terminology and ways in which results are displayed in student data reports (Lukin et al., 2004; Underwood et al., 2010; Zapata-Rivera & VanWinkle, 2010; Zwick et al., 2008). Common report formats communicate poorly and thus communicate misinformation because their creators do not know how to communicate intended messages (Few, 2008). For example, score reports for administrators are frequently not designed in ways that are easy for administrators to interpret (VanWinkle et al., 2011). Because there are many readers of reports, reports must include sufficient information to maximize the accuracy of their interpretations, explanations, clearer titles, and more guidance on where to read first, in a way that helps all users (Hattie, 2010).

Not a criticism of educators. These data analysis difficulties should not be mistaken as criticisms of educators, and the problem should not be mistaken as failure on the part of educators. Rather, evidence suggests educators represent highly skilled and intelligent individuals whose school districts are predominantly employing research-based recommendations to which they have access to improve data use.

For example, 99% of American teachers have bachelor's degrees, 48% have master's degrees, and over 7% have more advanced graduate degrees (Papay, Harvard

Graduate School of Education, 2007). In addition, evidence suggests educators are generally embracing data and technology use. For example, most educators are eager to analyze and then act on the data they see (Hattie, 2010; van der Meij, 2008), teachers indicated overwhelming support for using technology to improve learning, and 85% of teachers reported daily use of technology to support teaching (Bill and Melinda Gates Foundation, 2012). Furthermore, most schools and districts are already following recommendations within their control (e.g., PD and staff supports) to improve their data use. For example, districts devote between 1% and 8% of their operating budgets to providing professional learning (Killion & Hirsh, 2012), and 85% of principals indicate it is very important for them to be able to use student achievement data to improve instruction (Metropolitan Life Insurance Company, 2013). Likewise, 59% of the 211 participants in this *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study indicated they had underdone at least some PD in the past year devoted specifically to how to analyze student data.

This study was not based on the misconception data analysis errors are due to flaws in the educators who make them. Rather, it is based on recognition that a population surpassing the general public in schooling and intellect yet still struggling with data analyses, despite its own efforts to rectify the problem, might be using tools that are inherently flawed in their ability to render accurate analyses. Educators use data system-generated reports to make decisions that impact students (VanWinkle, Vezzu, & Zapata-Rivera, 2011). Data systems and their reports are the tools educators use to analyze data. Thus this study investigates the data system reporting environment.

In order to improve data use, practitioners and researchers need to gather empirical evidence to support different ways in which data is reported (Lyrén, 2009). Education stakeholders need to look for ways in which the data analysis tools educators use might be improved in order to better serve these educators and the students they work to help. This study contains evidence of specific ways current data systems are contributing to educators' failed data analyses and of specific ways these data systems can be improved to render more accurate analyses when used by educators.

Theoretical framework summary. Many educators struggle to understand how to translate data into specific actions (Cho & Wayman, 2009; Ingram, Louis, & Schroeder, 2004; Supovitz & Klein, 2003; Wayman & Cho, 2009). This problem persists despite educators' advanced academic backgrounds and efforts to improve their own data use. Data systems can provide solutions to the problem of educators' flawed data analyses, but they commonly do not (Marsh, Pane, & Hamilton, 2006).

The vast majority of stakeholders who need to use data to comprehend and raise student achievement are not trained statisticians, and they need additional information to teach them how to understand the data they view and how to use and apply the data to decision-making that can help students succeed (DQC, 2009). The researcher of this study sought to determine how data systems can provide this needed information to facilitate more accurate data-informed decision-making for the benefit of students impacted by those decisions.

Research Questions

Research questions were used to explore the impact of three variables on data analysis accuracy: (a) labeling in the form of brief, cautionary verbiage in report footers;

and (b) supplemental documentation in the form of report abstracts and (c) interpretation guides. All three of these data analysis supports hold potential to improve educators' data-based conclusions, yet their prospective impact had not yet been measured. Thus research questions Q2a, Q3a, and Q4a, with null and alternative hypotheses for each, were designed to measure the supports' precise impact on educators' data analysis accuracy. Research question Q1, with null and alternative hypotheses, was designed to measure the precise impact of these supports on educators' data analysis accuracy, in terms of exposure to or use of *any one* of the supports. Currently, educators make frequent analysis errors when drawing data-based conclusions. Educators then use those conclusions to shape decisions and actions that impact students. Thus these research questions, which were used to determine ways in which data systems can better facilitate accurate data analyses, hold potential to help researchers – and those with which they communicate – to help students.

In order to thoroughly adhere to the study's theoretical framework, research questions also addressed framing (see the *Chapter 2: Literature Review: Behavioral Economics and Data-Informed Decision-Making: Framing* section for an explanation of the term), which was another key reason behind the necessity of this study. Each of the three data analysis supports with which this study's research questions were concerned were framed in two different formats within handouts given to study participants. This was done because the best way in which to frame analysis support within a data system to specifically improve educators' analyses had not yet been determined. Suggested ways to present analysis guidance in footers, abstracts, and interpretation guides was utilized in this study, but the best manner in which to frame these resources had not yet been

determined in regards to direct impact on analysis accuracy. Thus research questions Q2b, Q3b, and Q4b, with null and alternative hypotheses for each, were designed to measure the precise impact the supports' framing has on educators' data analysis accuracy. Educators' likelihood of using each of these supports was also factored into the answer for each of these questions. Thus the study's spectrum of research questions fills a void in education field literature by containing evidence that can be used to identify not only whether – and to what extent – data systems can help increase data analysis accuracy by providing analysis support within data systems and their reports, but also *how* those supports can best be provided.

Additional variables that could possibly have impacted educators' likelihood of using the investigated supports and/or educators' data analyses were also examined to help better understand the implications of findings in regards to all the research questions discussed above. For example, higher need student populations are sometimes thought to prompt educators to use data more frequently and thus with more success. Thus variables concerning relevant school site demographics were addressed by research questions Q5a, Q5b, Q5c, Q5d, Q5e, and Q5f, all of which measure each group's data analysis accuracy. As another example, one might wonder if educator veteran status rendered some educators to more frequently use added guidance than other educators, or to analyze data with more success. Thus relevant participant characteristics were addressed by the researcher with research questions Q6a, Q6b, Q6c, Q6d, and Q6e, all of which measure each group's data analysis accuracy. Educators' likelihood of using the investigated supports was also factored into the answer for each question discussed above.

Q1. What impact does data analysis guidance accompanying a data system report in the form of footer, abstract, or interpretation guide have on how frequently educators draw accurate conclusions concerning student achievement data?

Q2a. What impact does a footer with analysis guidelines on a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?

Q2b. What impact does the manner in which a footer is framed, in terms of moderate differences in length and text color, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?

Q3a. What impact does providing a report abstract, such as a one-page reference sheet with report purpose and data use warnings specific to the report it accompanies, with a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?

Q3b. What impact does the manner in which an abstract is framed, in terms of moderate differences in density and header color, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?

Q4a. What impact does providing an interpretation guide, such as a two-sided reference sheet with analysis guidance and examples specific to the report it accompanies, with a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?

Q4b. What impact does the manner in which an interpretation guide is framed, in terms of moderate differences in length and information quantity, have on its ability to

impact the frequency with which educators draw accurate conclusions concerning student achievement data?

Q5a. What impact does an educator's school site level type (i.e., elementary or secondary) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q5b. What impact does an educator's school site level (i.e., elementary, middle/junior high, or high school) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q5c. What impact does an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q5d. What impact does an educator's school site English Learner (EL) population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q5e. What impact does an educator's school site Socioeconomically Disadvantaged population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q5f. What impact does an educators' school site Students with Disabilities population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q6a. What impact does an educator's veteran status have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q6b. What impact does an educator's current professional role (e.g., teacher, site/school administrator, etc.) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q6c. What impact does an educator's perception of his or her own data analysis proficiency impact the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q6d. What impact does an educator's professional development over the past year, devoted specifically to *how* to analyze student data, have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Q6e. What impact does the number of graduate-level educational measurement courses an educator has taken have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

Hypotheses

$H1_0$. The null hypothesis was that accompanying a report with a support containing analysis guidance in the form of footer, abstract, or interpretation guide would not have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

$H1_a$. The alternative hypothesis was that accompanying a report with a support containing analysis guidance in the form of footer, abstract, or interpretation guide would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

$H2a_0$. The null hypothesis was that accompanying a report with a supportive footer containing analysis guidance would not have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

$H2a_a$. The alternative hypothesis was that accompanying a report with a supportive footer would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

$H2b_0$. The null hypothesis was that the manner in which a footer was framed, in terms of moderate differences in length and text color, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.

$H2b_a$. The alternative hypothesis was that the manner in which a footer was framed, in terms of moderate differences in length and text color, would have an impact on the frequency of accurate conclusions educators drew concerning student achievement data.

$H3a_0$. The null hypothesis was that including a report abstract with a data system report would not have a positive impact on the frequency with which educators drew accurate conclusions concerning student achievement data.

$H3a_a$. The alternative hypothesis was that including a report abstract with a report would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

$H3b_0$. The null hypothesis was that the manner in which an abstract was framed, in terms of moderate differences in density and header color, would not have an impact

on the frequency with which educators drew accurate conclusions concerning student achievement data.

H3b$_a$. The alternative hypothesis was that the manner in which an abstract was framed, in terms of moderate differences in density and header color, would have an impact on the frequency of accurate conclusions educators drew concerning student achievement data.

H4a$_0$. The null hypothesis was that including an interpretation guide with a data system report would not have a positive impact on the frequency with which educators drew accurate conclusions concerning student achievement data.

H4a$_a$. The alternative hypothesis was that including an interpretation guide with a report would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

H4b$_0$. The null hypothesis was that the manner in which an interpretation guide was framed, in terms of moderate differences in length and information quantity, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.

H4b$_a$. The alternative hypothesis was that the manner in which an interpretation guide was framed, in terms of moderate differences in length and information quantity, would have an impact on the frequency of accurate conclusions educators drew concerning student achievement data.

H5a$_0$. The null hypothesis was that an educator's school site level type (i.e., elementary or secondary) would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

H5a$_a$. The alternative hypothesis was that an educator's school site level type (i.e., elementary or secondary) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

H5b$_0$. The null hypothesis was that an educator's school site level (i.e., elementary, middle/junior high, or high school) would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

H5b$_a$. The alternative hypothesis was that an educator's school site level (i.e., elementary, middle/junior high, or high school) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

H5c$_0$. The null hypothesis was that an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

H5c$_a$. The alternative hypothesis was that an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

H5d$_0$. The null hypothesis was that an educator's school site English Learner (EL) population would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

H5d$_a$. The alternative hypothesis was that an educator's school site English Learner (EL) population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H5e_0$. The null hypothesis was that an educator's school site Socioeconomically Disadvantaged population would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H5e_a$. The alternative hypothesis was that an educator's school site Socioeconomically Disadvantaged population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H5f_0$. The null hypothesis was that an educator's school site Students with Disabilities population would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H5f_a$. The alternative hypothesis was that an educator's school site Students with Disabilities population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6a_0$. The null hypothesis was that an educator's veteran status would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6a_a$. The alternative hypothesis was that an educator's veteran status would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6b_0$. The null hypothesis was that an educator's current professional role (e.g., teacher, site/school administrator, etc.) would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6b_a$. The alternative hypothesis was that an educator's current professional role (e.g., teacher, site/school administrator, etc.) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6c_0$. The null hypothesis was that an educator's perception of his or her own data analysis proficiency would be related to the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6c_a$. The alternative hypothesis was that an educator's perception of his or her own data analysis proficiency would not be related to the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6d_0$. The null hypothesis was that an educator's professional development over the past year, devoted specifically to how to analyze student data, would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6d_a$. The alternative hypothesis was that an educator's professional development over the past year, devoted specifically to how to analyze student data, would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6e_0$. The null hypothesis was that an educator's number of graduate-level educational measurement courses would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

$H6e_a$. The alternative hypothesis was that an educator's number of graduate-level educational measurement courses would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Nature of the Study

This experimental, quantitative study measured how effective three data analysis supports, which are found in some data systems and can be added to others, are in improving educators' data analysis accuracy: (a) labeling in the form of brief, cautionary verbiage in data system report footers; (b) supplemental documentation in the form of report abstracts that can be reached via link in a data system and can also be printed to accompany printed reports, and (c) supplemental documentation in the form of interpretation guides that can be reached via link in a data system and can also be printed to accompany printed reports. Participants answered survey questions regarding student data reports they received, which featured varying levels and forms of analysis guidance. In addition to establishing the data analysis accuracy rendered by educators using reports with no added supports, the survey was used to measure the specific impact the three above-listed variables (a-c) have on educators' data analysis accuracy.

The researcher employed a cross-sectional sampling procedure when incorporating responses from 211 educators of all school levels spanning transitional kindergarten (TK) through twelfth grade, at all veteran levels, working in varied roles, and at schools with a range of demographics. These educators were employed at nine schools, six school districts, six cities, and three counties in California. Conclusions did not rely on participants' preferences or perceived value of supports, but rather were based on how the supports impacted participants' answers to data analysis questions based on data system reports. The findings of this study can be used to identify whether, how, and to what extent data systems can help increase data analysis accuracy by providing

analysis support within data systems and their reports, and thus fill a void in education field literature.

Significance of the Study

The FDA directs the pharmaceutical industry to accompany over-the-counter medication with textual guidance regarding its use but to also provide solid evidence on how effective its labeling is in reducing errors; to proceed without such research is considered negligent (DeWalt, 2010). Despite the common use of data systems to generate reports, research on aspects of report format and system support that could enhance analysis accuracy is scarce (Goodman & Hambleton, 2004). Research that was devoted to data system and report format, including how effectively this format communicates data to users, focuses on participants' preferences and participants' perceived value of supports. However, user preference can be the opposite of the reporting format that actually renders the more accurate interpretation (Hattie, 2010).

This study was used to examine how effective varied analysis supports are in improving data analysis accuracy, and it did not rely on participants' preferences or perceived value of supports. The findings of this study fill a void in education field literature by containing evidence that can be used to identify whether, how, and to what extent data systems can help increase data analysis accuracy by providing analysis support within data systems and their reports. Improvements data system and report providers make in light of this study have the potential to improve the accuracy with which educators analyze the data generated by their data systems. More accurate data analyses will likely result in more accurate data-informed decision-making for the benefit of students.

Definition of Key Terms

Abstract. Please see *Report Abstract*.

Accountability. State and federal accountability systems aim to improve student performance by pairing academic goals and standards with incentives for schools and districts, and they have increased in importance and pressure since the 2001 No Child Left Behind (NCLB) Act and the 2002 Elementary and Secondary Education Act (ESEA) (Gross & Goertz, 2005).

Achievement. Achievement constitutes what an individual has learned; in education, this refers to what a student has learned in school (Airasian, 2000).

Assessment. Assessment describes the process of using information about students and instruction to assist making decisions in and about the classroom (Airasian, 2000).

Mean/Average Percent Correct. The mean/average percent correct is acquired by totaling the number of questions all students with valid scores answered correctly for a test or reporting cluster, also called their raw scores, then dividing that number by the number of students with valid scores, then dividing that number by the total number of test questions for the test or reporting cluster, and then multiplying that number by 100 (California Department of Education, 2011).

California Standards Test (CST). The CST measures student performance on California content standards and constitutes the largest component of California's Standardized Testing and Reporting (STAR) Program (California Department of Education, 2011).

Choice Architecture. Choice Architecture refers to the organization of the context within which people make decisions, which greatly impacts decision-making (Thaler & Sunstein, 2008).

Dashboard. A dashboard is a visual and typically graphical display of the most important info a user needs, arranged on a single screen so it can be viewed at a glance (Few, 2006).

Data Literacy. Data literacy refers to one's ability to understand data and ask appropriate questions in relation to it (U.S. Department of Education Institute of Education Sciences National Center for Education Evaluation and Regional Assistance [USDEIESNCEE], 2009).

Data Mining. In education, data mining involves developing tools to discover patterns in education data, such as the learning of one-digit multiplication, in order to make predictions and appropriate plans of action (U.S. Department of Education Office of Educational Technology, 2012).

Data System. A data system is a computer system that aims to provide educators with student data to help solve educational problems (Wayman, 2005). Decision support systems (DSSs), data warehouses, and data marts can all be data systems, though these three systems each differ from one another (NFES, 2006). Other examples of data systems include student information systems (SISs), assessment systems, instructional management systems (IMSs), and data-warehousing systems, but distinctions between different types of data systems are blurring as these separate systems begin to serve more of the same functions (Bill and Melinda Gates Foundation, 2007).

Data Teams. A data team is a school-based group of educators who analyze data together to help one another in the data's analysis and use (USDEIESNCEE, 2009).

Data-Informed Decision-Making. Data-informed decision-making refers to the collection and analysis of data to guide decisions that improve student success (USDEIESNCEE, 2009). While data-driven decision-making is a more common term, data-informed decision-making is a preferable term since decisions should not be based solely on quantitative data (Knapp, Swinnerton, Copland, & Monpas-Hubar, 2006; USDEOPEPD, 2009).

Education. Education describes the institution or process designed to positively impact students in specific ways (Airasian, 2000).

G*Power 3. G*Power 3 programs facilitate statistical power analyses conducted in varied scientific fields (Faul, Erdfelder, Lang, & Buchner, 2007).

Help System. A computer-based help system stores supporting information, facilitates the search for supporting information, and retrieves information appropriate to each situation (Inoue & Tagawa, 2006).

Interpretation Guides. Also called *interpretive* guides, interpretation guides accompany some reports to answer questions users might have concerning the reports, such as by explaining the test purpose, term definitions, scoring guides, how to read the report, etc. (Goodman & Hambleton, 2004). An interpretation guide can also be thought of – and called – a reference guide.

Learning Analytics. Learning analytics involves applying tools to discover patterns in education data, such as in classrooms or schools, in order to make predictions

and appropriate plans of action (U.S. Department of Education Office of Educational Technology, 2012).

Longitudinal Data. Longitudinal student data is collected over time to facilitate more thorough analyses of student performance than partial histories would offer (USDEOPEPD, 2010).

Measurement. Measurement is defined as the process of assigning numbers or categories to performance based on specific rules and standards (Airasian, 2000). When teachers' understanding of data use concepts was studied, 80% demonstrated an understanding of measurement error, and 37% an understanding of multiple measures (though only 1 case study teacher spoke explicitly of the need for multiple measures) (USDEOPEPD, 2011).

NAEP. Mandated by Congress in 1969, the National Assessment of Educational Progress (NAEP) has been used to monitor student achievement countrywide (NRC, 2001).

No Child Left Behind (NCLB) Act of 2001. NCLB carried a federal mandate for schools, districts, and states to raise and report on student performance and called attention to the potential of data use to improve student achievement (Wayman, 2005).

Percent Correct. Percent correct is acquired by dividing the total number of questions answered correctly for a test or reporting cluster, also called the raw score, then dividing that number by the total number of test questions for the test or reporting cluster, and then multiplying that number by 100 (California Department of Education, 2011).

Performance Levels. Every scale score on a California state assessment translates to one of five performance levels (California Department of Education. 2011).

Professional Development (PD). Successful professional development (PD) provides educators with a continual process of learning and improvement through multiple strategies, ranging from online training to traditional workshops (Southern Regional Education Board, 2009).

Report. Please see *Score Report*.

Report Abstract. Also called a summary, an abstract summarizes a report's main points, clarifies the report's scope, and serves as an advance organizer, helping the user to mentally structure the report's many details (Aschbacher & Herman, 1991). The abstract provides supplemental information such as the report's description, purpose, intended audience, content, format, and cautionary information concerning data misconceptions and use (Illuminate Education, 2012). A report abstract can also be thought of – and called – a reference sheet.

Scale Score. A scale score is a raw score that has been altered/scaled to account for differing difficulties from one administration year to the next so that performance from different years on the same test may be compared, as percent correct and raw scores do not allow for this (California Department of Education. 2011).

Score Report. Score reports communicate test results to stakeholders in a variety of ways that should be easy to use and understand (De Jong & Zheng, 2011). Graphics are a recommended component for score reports (Hattie, 2010; NRC, 2001; VanWinkle et al., 2011).

Standardized Assessment. Standardized assessment is a test that is administered and scored in the same manner for all students taking the test, and thus all students' results on the test may be interpreted in the same way (Airasian, 2000).

Standardized Testing and Reporting (STAR) Program. California's STAR Program consists of four components for students in grades 2-11: California Standards Tests (CSTs) for students without disabilities, California Modified Assessment (CMA) for students with disabilities but not severe cognitive impairments, California Alternate Performance Assessment (CAPA) for students with severe cognitive impairments, and Standards-based Tests in Spanish (STS) for some English Learners (ELs) (California Department of Education, 2011).

Test. Like an assessment, a test is a methodical procedure for obtaining a sample of student performance (Airasian, 2000).

User. Underwood, Zapata-Rivera, and VanWinkle, (2008) noted users of data systems vary in needs; for example, teachers might be "novice users" (p. i) who lack computer experience, "tech-ready users" (p. i) with more computer experience or a desire to investigate data, or "tech-savvy users" (p. i) wanting to access all data system features.

Summary

While a doctor isn't present to explain an over-the-counter medication's use, medicine bought in a store comes with a detailed label outlining its purpose, ingredients, dosage instructions, and dangers, as well as supplemental documentation offering more room to expound upon the contents' recommended use. It would be negligent to sell medicine without such guidance on how to use it wisely, as this would risk the lives of those the medicine is used to treat (Brown-Brumfield & DeLeon, 2010, DeWalt, 2010). Meanwhile, educators are using data to treat students, yet they are operating without the data-equivalent to over-the-counter medicine: reports generated in data systems typically contain insufficient labeling and documentation to guide users in the data's use. The vast

majority of stakeholders who use student data are not trained statisticians, and they need the data they view to be accompanied with additional information to teach them how to understand and use the data (DQC, 2009). Yet educators are using data systems and data system reports that do not feature guidance on the data's appropriate use, and they're using it to inform decisions that impact students – much like ingesting medicine from an unmarked or marginally marked container or using such medicine to blindly treat the wellbeing of a child. Hampton (2007), Qin et al. (2011), and Clay (2012) offered or called for label recommendations similar to those recommended by the FDA for over-the-counter medication labels. Label conventions can result in improved understanding on non-medication products, as well, if they are included (Hampton, 2007; Qin et al., 2011).

Educators are in dire need of analysis help. There is strong evidence many users of data system reports have trouble understanding the data (Hattie, 2010; NRC, 2001; Wayman et al., 2010; Zwick et al., 2008). For example, in a national study of districts known for strong data use, only 48% of teachers correctly interpreted data (USDEOPEPD, 2009). It is unlikely teachers at districts where data use is less emphasized would make more accurate data analyses than those described in a study of districts considered exemplars of data use (USDEOPEPD, 2011).

Research contains evidence that while PD and staff supports are beneficial to improving data use, these approaches are not without limitations, and they are not enough. In addition, the typical educator is analyzing data while alone, unaccompanied by a data expert when he or she makes data-driven decisions (USDEOPEPD, 2009). When these analyses take place, the educator is merely accompanied by the data system

and/or its reports. Even when analyses benefit from PD and staff supports, students deserve for educators to use *all* possible supports for improved analysis accuracy.

Research on aspects of report format and system support that can improve analysis accuracy is scarce (Goodman & Hambleton, 2004). Research that was devoted to data system and report format focuses on participants' preferences and participants' perceived value of supports as opposed to measuring supports' actual impact on interpretation. This study was used to examine exactly how effective varied analysis supports are in improving data analysis accuracy. The findings of this study contribute to literature in the field by helping to identify how data systems can help increase data analysis accuracy by providing analysis support within data systems and their reports. Due to the impact educators' data analyses have on students, this means the findings have the potential to benefit students. It is the strong conviction of this researcher that students deserve for educators to use *all* possible supports for improved analysis accuracy in an effort to completely eliminate – rather than merely reduce – their data analysis errors when using those analyses to make decisions that impact students' lives.

Chapter 2: Literature Review

The purpose of this quantitative study was to investigate the degree to which including different forms of data usage guidance within a data system can improve educators' understanding of the data content, much like including different forms of usage guidance with over-the-counter medication is needed to properly communicate how to use its contents. The literature review required the investigation of numerous topics related to the topic, as it incorporated research into data use, tools, and practices; data analysis accuracy problems; possible solutions to these problems; and limitations to solutions to these problems. Non-education topics like over-the-counter medication labeling and report design, which is not exclusive to education, were also explored.

Because this study involves an analogy between over-the-counter medication labeling and similar labeling typically missing from student data systems and their reports, both education and medical research were explored. Printed publications, such as books and journals cited in this paper's reference list, were utilized in addition to sources accessible through online searches. Education related topics involved the reading of literature tied to keywords such as *analysis errors*, *analysis support*, *data analysis*, *data and assessment management*, *data-driven*, *data error*s, *data-informed*, *data management*, *data use*, *data system*, *footers*, *interpretive guide*, *interpretation guide*, *report abstract*, *report design*, *report format*, and *report use*, as well as variations of these terms. The terms related to report design and use, which are not exclusive to education, were also explored in a non-education context. Terms related to decision-making and not exclusive to education, such as *behavioral economics*, were also incorporated into the review. Solutions for improved data use that arose – such as *data dialogue*, *data discussions*, *data*

team, *data expert*, *instructional coach*, *leadership*, *PD*, and *Professional Learning Community (PLC)* – were further explored to determine the approaches' comprehension and success. Medical and pharmaceutical topics involved the reading of literature tied to keywords such as *directions*, *Food and Drug Administration requirement*, *instructions*, *labeling*, *labels*, *over-the-counter*, *medication*, *medicine*, and *safety*, as well as variations of these terms. The literature search strategy involved the use of numerous research databases and mainly included:

- California Teachers Association (CTA) California Educator Archives (http://legacy.cta.org/media/publications/educator/archives/California+Educator+Archives.htm)
- Center on Education Policy (www.cep-dc.org)
- Ebrary (www.ebrary.com)
- EBSCOhost (www.ebscohost.com), which includes MEDLINE
- Education and Information Technology Digital Library (EdITLib) (www.editlib.org)
- Education Resources Information Center (ERIC) (www.eric.ed.gov)
- GALE CENGAGE Learning Academic OneFile (www.gale.cengage.com/PeriodicalSolutions/academicOnefile.htm)
- Google Scholar (http://scholar.google.com)
- Institute of Education Sciences (IES): National Center for Education Statistics (NCES) Publications and Products (www.nces.ed.gov/pubsearch)
- The Journal of the American Medical Association (JAMA) (http://jama.jamanetwork.com/journal.aspx)

- National Association of Elementary School Principals (NAESP) Archives (http://www.naesp.org/principal-archives)
- National Association of Secondary School Principals (NASSP) Knowledge Center (http://www.nassp.org/knowledge-center)
- Online Educational Research Journal (OERJ) http://www.oerj.org
- ProQuest (www.proquest.com)
- Public Impact (www.publicimpact.com)
- RefWorks (www.refworks.com)
- Sage Journals (www.sagepub.com/journals.nav)
- Sage Reference (www.sage-ereference.com)
- Teachers College Record (www.tcrecord.org)
- University of Texas at Austin: Department of Educational Administration, College of Education Data Use Publications (http://edadmin.edb.utexas.edu/datause/publications.htm)
- U.S. Department of Education (http://find.ed.gov)
- Wiley Online Library, which includes British Journal of Education Technology (BJET) (http://onlinelibrary.wiley.com)

This study's researcher also initiated dialogue with researchers of previous, relevant studies. For example, in 2011 she attended a one-day Presenting Data and Information course at the Westin San Francisco Market Street by Edward Tufte. In 2012 this study's researcher visited with Dr. Jeffrey C. Wayman at his workplace at the University of Texas at Austin, attended one of his courses, and discussed his and her research. In 2013 this study's researcher conversed with Faris M. Sabbah via email

concerning his research through San Francisco State University; California State University, East Bay; and San Jose State University, as well as hers. Also in 2013, this study's researcher attended an interactive, online interview with John Hattie (*John Hattie on What Actually Works in Schools to Improve Learning*) conducted by Steve Hargadon through the Future of Education (Hattie & Hargadon, 2013).

The literature review begins with an introduction that provides background information concerning the topic as it relates to research and also introduces the main literature review findings. Next, a more extensive history is given of research on data systems, their reports, their use, and findings concerning the most effective ways of improving the accuracy of educators' data analyses. A section is then devoted to national reporting standards that feature specific recommendations for how educational data should be displayed and communicated, as this constitutes an important piece of the topic's history. The current state of educators' data analysis skills are then explored in relation to the literature, as this illustrates the data analysis accuracy problem this study was to help solve. Controversy concerning the best way to improve data analysis accuracy is then detailed, focusing on the two theories that dominate most literature on the topic: PD and staff supports, with the latter including such resources as site leaders, data teams, data experts, and/or instructional coaches. However, these supports outside of the data systems are not enough, as the next section highlights. The literature review builds to three questions that remain unanswered by current research, and each has its own section. Unanswered Question 1 covers *content*, as there are conflicting findings concerning what additional analysis information should be included with data reports. Unanswered Question 2 covers *quantity*, as research contains evidence not all

recommendations should be included since the magnitude can overwhelm educators and lead to less success than the inclusion of fewer details. Unanswered Question covers the *impact* of each component on analysis accuracy, since not every recommendation may be accommodated, and research is needed to determine how likely each data system support is to increase analysis accuracy. Finally, this literature review concludes with a summary of findings and key points that are essential to this study.

Introduction

Recommendations concerning the best ways to display data have been around for many years, with William Playfair's work from the 1700s and 1800s being the most influential (Wainer, 1992). Large-scale assessments have been a component of U.S. education since the 1800s and have been widespread since the 1920s (Hamilton & Koretz, 2002). Thus research on the best ways to present assessment data so as to improve analysis preceded the rise of student data systems, which accompanied and grew with the Internet's appearance in school districts in the 1980s (see the *Chapter 2: Literature Review: History of Specific Research Contributions* section for a historic timeline of research contributions on reporting problems and recommendations for improving report design). The NCLB Act of 2001 increased pressure on educators to raise student achievement, which increased the demand for data systems that facilitate analysis of educational data (USDEOPEPD, 2009). NCLB led to reports for multiple subjects being distributed at the state, district, school, subgroup, and student levels for parents and teachers of 22 million students per year, yet the reports are not in accordance with any nationally recognized reporting standards (Hattie, 2010).

The literature contained evidence of controversy concerning the best way to improve data analysis accuracy, but also evidence that supports outside data systems are not enough and more research is needed to improve data systems and their reports. More research needs to be done to specifically investigate how reports can better facilitate correct interpretations by its users (Hattie, 2010). Even within data system and report research, findings offered controversy concerning the recommended content and quantity of analysis support. Findings also fell short of concrete evidence for best practices, as studies adhered to the common approach of examining educator reporting *preferences* as opposed to measuring reporting options' impact on analysis accuracy. Thus findings were inadequate in some areas and conflicting in others. Appropriately, the literature also specifically states more research is needed on the topics of better communicating student data such as test results, and on improving educators' data analysis accuracy. The *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study addressed both of those research topics within the context of analysis guidance that data systems and their reports can provide.

History of Specific Research Contributions

Although the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study concerned how data is generated in online data systems, which accompanied the Internet's increasing appearance in school districts in the 1980s, the type of reports these systems generate typically involve a traditional, printable report format. Thus research concerning ideal report format and data delivery that predate this technological age can be applicable. However, such a history could date back to at least the 1800s and comprise a thick book. Thus this literature history will include the most

influential research publications from the last two decades. Also, the masses began using personal computers around 1990 (Leeson, 2006). Since online student data systems are used on computers and accessed from classrooms, offices, and homes, this makes the 1990s an even more appropriate time period with which to begin this literature review's history.

The majority of literature on educators' data analysis accuracy and related topics does not include mention of the role in which the design and/or features of data systems and their reports can impact that accuracy. Thus specific bodies of research stand out as key milestones in the evolution of research on the topic. This section is meant to profile the most important publications in the field in regards to the study's topic while providing a sense of how these contributions built upon one another over time. Thus paragraphs in this section only (below) are devoted mainly to specific publications. However, this section is only one of 12 sections in the literature review, and the remaining 11 sections of the literature review organize assorted research sources around specific subtopics such as themes and unanswered questions, as opposed to adhering to a timeline format.

1991. *Aschbacher and Herman*. By the early 1990s there was little research on the data system's impact on data analysis accuracy, yet some attention was given to the types of reports data systems generate. For example, prompted by research on the common failures of assessment reports and by the lack of research on the manner in which results are presented to intended audiences, a study of reporting practices in 30 U.S. states resulted in reporting guidelines such as including information directly on graphs and tables rather than requiring users to look elsewhere for help, as well as the use of footnotes and explanations (Aschbacher & Herman, 1991). The research team

compiled a list of report details that could help educators use data and avoid errors when doing so. These included considering the report's audience and purpose; using summaries or abstracts; "chunking" information in nine or fewer categories; being comprehensive and balanced; capturing and focusing the user's attention with color-coding, graphs, or question and answer headings; including explanatory footnotes; avoiding excessive negative wording; using a format appropriate for the report's purpose; pairing data with words since users more easily recall words than numbers; using consistent displays in cases where it is appropriately-suited to the data well; including titles, headings, and footnotes on graphical and tabular displays; keeping titles concise but unique; labeling information directly on charts whenever legends can be avoided; and selecting some graph types over others due to ease of use. Aschbacher and Herman (1991) also recommended providing all information needed for an analysis on one page, arguing that placing information on separate pages will interfere with users' ability to make connections between the data and information contained on them. Other research professes the benefits of separate-page abstracts and guides. This discrepancy thus constitutes another controversy that findings from the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study help to answer.

1992. *Wainer*. The American Educational Research Association (AERA) profiled how graphs and tables are interpreted and how tables should be displayed, and noted the easiest and most common way to test graph readability is to use elementary level questions that can be answered through data extraction (Wainer, 1992). AERA recommended graphs be included in reports to answer questions involving data extraction, trends, comparisons, and groupings, whereas tables should be used to

communicate data in a logical order, with rounded numbers in almost all cases for ease of use, and with summaries of rows and points to serve as crucial comparisons to other data in the table (Wainer, 1992). Though it was part of a piece on understanding graphs and tables, AERA's contribution was important in its acknowledgement that the way data is displayed – rather than merely the educator viewing it – influences how the data is analyzed. Wainer later demonstrated the importance of this concept by writing a book on the same graphical topic (Wainer, 1997).

1993-1996. *Hambleton and Slater*. Prompted by research by Jaeger (1992), Linn and Dunbar (1992), and Koretz and Deibert (1993) indicating National Assessment of Education Progress (NAEP) reports were being misinterpreted, interviews with 59 educators and policymakers who expressed medium to high interest in national student achievement drew important lines between educator analysis errors and report design (Hambleton & Slater, 1996). Participants included 12 state education agency administrators, 17 Department of Education consultants and researchers, two education reporters, eight school administrators, seven legislators and related staff, and 13 national and regional education organization directors and assistants. Many of the interviewees demonstrated limited statistical knowledge and analysis errors and misunderstandings concerning the data were common, prompting a recommendation to reduce obstacles by field testing data displays, simplifying reports, and making each report easier for its intended audience to understand. Additional recommendations included avoiding overly complex tables, indicating cases where performance bands have been added together, including descriptive information and clear examples, providing users with an example of how to read each chart, making color differences clear, explaining bands within bars, and

only including legends when necessary. An encouraging finding was that when a report was explained, a different yet similar report was more easily understood than if the user had never had the initial report explained to him or her.

Despite participants' analysis difficulties, nearly all understood the data once it was explained to them, prompting a recommendation that an example of how to read each chart be included in addition to the directions already present above charts. Some interviewees referred to footers for explanations; however, they helped little due to statistical jargon, prompting the recommendation that reports be understandable without reference to text (Hambleton & Slater, 1996). This last finding is not without controversy in educational research, as not every piece of assessment data is simple enough to speak for itself and other researchers recommend footers, as detailed in this literature review.

1997. *Tufte*. The author and speaker Edward Tufte emerged as a leading expert in data visualization, applying to report design as it influences the communication of data and information. In fact, some refer to Tufte as the "godfather of modern data visualization" (Schwabish & Schultz, 2013). Tufte (1997, 2001, 2006) asserted that poor displays interfere with users' ability to read graphical displays. Other recommendations included providing details but keeping them concise, communicating both general and specific messages, featuring the actual numbers directly on graphs, making large data sets easy to comprehend, and using proximity to facilitate appropriate comparisons of disparate data sets. While Tufte regularly emphasized the pitfalls of clutter, he also recommended maximizing data density by presenting many numbers in a small space, and he encouraged report designers to do 'whatever it takes' to communicate intended messages (Tufte, 2011). He balanced and applied other conflicting ideas, as well,

demonstrating the flexibility needed in the ‘whatever it takes’ approach. Not exclusive to the field of education, Tufte (1997, 2001, 2006, 2011) contradicted other research in his encouragement of using uncommon displays rather than sticking exclusively to common table and graph types, which other research contains evidence educators have more success in correctly reading.

1999. *Zenisky, Hambleton, & Sireci*. A report on how National Assessment of Educational Progress (NAEP) results are communicated in the Internet age named score reporting as the most challenging aspect facing test agencies, as opposed to merely crafting a quality test (Zenisky, Hambleton, & Sireci, 1999). NAEP reports have been a popular focus in education report design research, since NAEP is the national assessment system most U.S. schools administer at some point in time. A changing selection of schools from districts are selected for NAEP participation each year, and only some grades levels are selected to test, so there is not administration continuity from the schools’ standpoint; however, due to national exposure to the test, educators from different states are more likely to share familiarity with NAEP results than they are with results from state or local assessments (Gorman & National Center for Education Statistics [NCES], 2010). However, NAEP results are not available for individual students, classes, or school sites (Gorman & NCES, 2010), and educators’ limited and varying exposure to NAEP results leaves them less accustomed to working with NAEP data than that from their state or local assessments. Using NAEP results for a national study concerning data reporting is thus advantageous, whereas using state assessment results would be more advantageous for a study conducted in a single state.

Using NAEP reporting as an example, the literature called for the inclusion of context when reporting a score earned, such as group comparisons or descriptions of strengths, and noted the shift to online reporting as a chance to fundamentally change how student data is communicated (Zenisky et al., 1999). Like Hambleton and Slater (2006), the literature called for student data to be accompanied by information specific to the audience the report is meant to target, such as educators, parents, or the media. The report also cited the power of giving users interactive, web-based, creative, and innovative tools to accompany data reports, such as multimedia and clickable data, the ability to manipulate the format of tables or graphs, the ability to manipulate information such as score type, and the ability to manipulate result types such as aggregate level or gaps between subgroups. However, not everything online must be interactive, as there is value in downloadable files that present information in easy-to-print formats (Zenisky et al., 1999). This is an important point, as data analysis is often done with a data system's reports but without all stakeholders actually using the data system online. For example, while some teachers (44%) use the data system directly, others (56%) have access but do not use the data system directly and instead only read printed versions of reports others used the data system to generate (Underwood et al., 2008). There is even evidence that such a practice is recommended for some users. Viewing a data system's report on the computer versus printed can negatively impact how it is interpreted; for example, someone who correctly interprets a printed report can make mistakes when scrolling is involved, users are more likely to scan a report on a computer that they would read carefully when printed, and users' inability to mark on the screen can reduce the credibility users attribute to reports (Hattie, 2010; Leeson, 2006).

The NAEP studies involved researchers asking varied NAEP audiences to comment on their reporting interests and preferences (Zenisky et al., 1999). This has been the prevailing approach in data system and report design research, where studies have examined which formats users *prefer* rather than which formats are shown to increase users' accuracy when analyzing data contained in data systems and their reports. Another research discrepancy is that not all studies promoted the use of interpretation guides and footers. For example, Harris (1999) was extensive in his coverage of information graphics available, how to use them, how to design them, and how to interpret them, yet he did not address such aspects as interpretation guides or footers.

2001. *The National Research Council (NRC)*. An examination of NAEP reporting practices noted attention to reporting formats could become more urgent as educators and non-educators of varying statistical sophistication strive to understand scores, and it acknowledged the error rate of even those who carefully study data reports, citing main problems with data reports as: high-level of statistical knowledge is assumed, information overload and report density, attempts at redesign increase clutter, infrequent use of graphics, and reports require unnecessary mental calculations (The National Research Council [NRC], 2001). Recommendations included avoiding too many technical terms, concepts, and symbols; using white space and variation to make reports appear easy-to-understand; avoiding three-dimensional visual displays; providing calculations so no mental arithmetic would need to be performed by the user; and favoring graphs over tables unless tables would result in more accurate interpretations of the data. The NRC (2001) made an important concession in acknowledging that the goal of a report should never be compromised in an effort to make the data appear less

intimidating or more accessible. That concession can be applied when report designers seek to achieve the balance discussed in the *Unanswered Question 1: Content*; *Unanswered Question 2: Quantity*; and *Unanswered Question 3: Impact of Each Component on Analysis Accuracy* sections of this literature review.

2002. *Fast and the State Collaborative on Assessment and Student Standards (SCASS) Accountability Systems and Reporting (ASR) Consortium*. The NRC's findings were echoed by that of Hamilton and Koretz (2002), who determined reporting format impacts how useful data is to stakeholders. The research team found that assessment results must be reported in an accessible manner for test-based accountability to work. However, their exploration of reporting types had to do with measurements – such as norm-referenced versus criterion-referenced – and related test format rather than supports. The SCASSASR Consortium, was also concerned with reporting as it relates to accountability, except it dealt more specifically with report format when it released *A Guide to Effective Accountability Reporting* (Fast & State Collaborative on Assessment and Student Standards Accountability Systems and Reporting Consortium [SCASSASR], 2002). SCASSASR noted accountability reports should contain adequate interpretive information, including cautions concerning possible misinterpretations, and should be designed with the goal that even one's next-door-neighbor should understand their meaning. Fast and the SCASSASR Consortium's recommendations served as clear support for the variables selected for the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study, which was used to identify the specific value of each variable.

2003. *Jaeger*. Commissioned by the NAEP Validity Studies (NVS) Panel and promoted by the U. S. Department of Education Institute of Education Sciences, Jaeger (2003) built on the work of the NRC (2001). While offering similar recommendations for reporting, he included details on how the history of NAEP reporting and research led to proposed changes (Jaeger, 2003).

2004. *Goodman and Hambleton*. Like Hambleton (2002) and the NRC (2001), Goodman and Hambleton (2004) acknowledged there was clear evidence many users of assessment reports had trouble understanding and interpreting the data they contain. The research team examined the student score reports and related interpretation guides of three United States (U.S.) commercial testing companies, 14 U.S. states' Departments of Education, and two Canadian provinces' Departments of Education. In one of the most oft-cited works on report design, the research team declared that while much attention has been devoted to the quality of assessments, very little attention or research has been concerned with ways in which the assessment results are reported and used. There is a clear need for research identifying how assessment results can most effectively be reported.

After exploring factors contributing to difficulties when trying to understand large-scale test results, the research team released recommendations for reporting student-level results, but the impact of each recommendation was not specified. Including text can improve chart and table interpretation, and including a glossary of terms can help to more effectively report results. Recommendations included making reports uncluttered and attractive, designing reports with a manageable number of purposes and desired interpretations in mind, making color use purposeful, including contextual information

with jargon-free language, and adding an interpretation guide to accompany each assessment score report. A large number of teachers misunderstand some types of information, and interpretive information reduces many of the difficulties they have in forming accurate interpretations. Goodman and Hambleton's (2004) recommendations served as clear support for the variables selected for the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study and for the way in which its report handouts were constructed.

2005. *Light, Wexler, and Heinz.* Research conducted by the Education Development Center's Center for Children and Technology (Light, Wexler, & Heinz, 2005) examined what it referred to as the "three dimensions" involved in transforming data into knowledge for educators: how data becomes usable information, how the data system used impacts this process, and how educators' prior knowledge impacts the process. The report was noteworthy in documenting that a decision support systems' design impacts how users turn data into knowledge. To truly be a decision support system, a data system needs robust reporting tools that can include explanatory information within charts, legends, citations, explanations, and other information to clarify the data's meaning (NFES, 2006). Light, Wexler, and Heinz (2005) suggested the systems include explanations and background information within the tool itself in an effort to support users' ability to understand the data without requiring additional, outside tools. These suggestions complimented other researchers' assertions that including supporting text such as footers, abstracts, and interpretation guides could have the potential to increase users' understanding of the data that the data system is being used to display. Prior to the *Over-the-Counter Data's Impact on Educators' Data Analysis*

Accuracy study, an exact measurement of each support's potential had not been determined.

***Benadom*.** Around the same time, Benadom (2005) touted the benefits of using a data system that puts data in a more comprehensible format as it stores it, generates reports and parent letters, and prompts changes in staff development. Benedom (2005) reported that an improved format would allow teachers, parents, and administrators to better help students, as they could more quickly understand and act on the results. These assertions were substantiated by Poger and Bailie (2006), who also wrote that reports need to be easy to interpret.

2006. *Rennie Center for Education Research and Policy*. Rennie Center for Education Research and Policy (2006) issued a comprehensive policy brief concerning tools and trends in data-informed teaching in which it acknowledged technology-related problems could impede teachers' ability to analyze test data properly. The brief noted teachers have very little time for data analysis, and this problem worsens as assessment frequency and complexity increases. It also noted teachers are far more likely to use data if it is presented in a user-friendly format. This research team constituted one of the few voices noting limitations in the most popular supports for data analysis, stating translating data into action is complex, and in order to effectively use data analysis tools teachers will need ongoing support; these are offered in the form of coaches and PD, but at a cost. Data systems do not typically include proper support for interpreting data and turning results into action, despite the fact that teachers do not often know how to translate data into action, and this could become the biggest challenge facing effective data use once educators are accessing technology otherwise deemed adequate. Teachers need to be able

to understand the data in the format a data system provides without any formal knowledge of statistics, or else they are not likely to use the system (Rennie Center for Education Research and Policy, 2006). These findings served as clear support for the variables selected for the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study, as well as the problems that inspired this study.

Marsh. Having data does not mean it will be used effectively (Marsh et al., 2006). Marsh, Pane, and Hamilton (2006) echoed previous researchers' findings concerning educators' data analysis struggles, stating not all educators have the skills and time needed to successfully use data to inform decisions, and educators' incomplete understanding of statistics can lead them to draw false conclusions from data. School staff often lacks the ability to interpret data. Educators often lack data analysis skills and the support needed to translate data into next steps; solutions include increased staffing assignments but also utilizing user-friendly data systems that provide options for analyzing data. This research team also acknowledged the gravity of this epidemic when they noted more research into educators' faulty analyses and misuse of data is needed.

Marsh et al. (2006) supported popular approaches to improved data use, stating PD and support from data expert staff can improve data use. However, their writing stood apart from most literature recommending such supports as it also acknowledged limitations. The most common method of supporting data-informed decision-making is PD focused on understanding test data, but its value varies, the majority of teachers and principals do not find it to be helpful, and sessions do not typically cover how to use test results for instructional planning. Site leaders are another source of data analysis support, but the quality of leadership varies. A report prepared for IES by Regional Educational

Laboratory Midwest confirmed the limitations of the most popular data analysis supports in school districts, noting data staff and training resources can be limited at the local level, as is staff with proper data analysis experience and skills at the state level (McDonald, Andal, Brown, & Schneider, 2007). Marsh's work promoting popular data analysis supports in the form of instructional coaches on was continued by Marsh, McCombs, and Martorell, F. (2010).

Marsh et al. (2006) explored solutions that research literature has largely ignored. More research is needed to help practitioners understand the best ways to present data and help staff translate data into information for improved instruction; for example, researchers could improve displays so that educators can more easily identify trends, regardless of their statistical backgrounds. Data systems can provide solutions, but they commonly do not. Turning to technology for support with data analysis is less common than turning to leaders or PD, and research contains evidence the majority of teachers do not find this to be a helpful means of support, perhaps since their data systems lack key components such as easy access to multiple data sources. Online data systems reduce time needed to generate data reports, but they still require educators' time in order to know how to act on the data, and lack of such time is limiting the data's use at many sites, meaning that few sites offer this critical component of data-informed decision-making (Ingram et al., 2004; Marsh et al., 2006). These assertions are highly refreshing in that they are part of rare acknowledgement in the research community that data systems should do more to support accurate data analyses.

The literature also supported the theory that data system report studies should not necessarily involve the direct use of data systems. Data accuracy, data access, technical

support, and training can all affect an educator's ability to understand and use data (Marsh et al., 2006). Without proper access and possibly technical assistance, educators can misinterpret data. Marsh et al.'s (2006) findings influenced the manner in which respondents' data analyses were facilitated in *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study.

2007. ***Minnici and Hill.*** Noting data interpretation has become increasingly vital to school reform, Minnici and Hill (2007) explored state-level system limitations in a report by the Center on Education Policy (CEP). The research team examined state education agencies' ability to provide NCLB-compliant accountability systems and analyzed the annual CEP survey data of officials in all 50 U.S. states, as well as interviews with 15 prominent state education officials from 11 U.S. states. The study indicated that while state's progress in offering data systems is being tracked, state-level staff are also at varying stages of being able to actually analyze the data these systems generate. Some state educators see their shift from mere data warehouses to dynamic data systems that facilitate data-informed decision-making at the state and local levels as the biggest change in their data systems (Minnici & Hill, 2007).

Perie, Park, & Klau. The Council of Chief State School Officers Accountability Systems and Reporting State Collaborative commissioned a paper outlining recommended components for educational accountability models, which covered the communication of data through reporting (Perie, Park, & Klau, 2007). The checklist for communicating accountability results recommended that reports communicate all relevant data clearly, promote accurate interpretation and use of data, use a format that helps schools learn how to use the data, apply the latest research in effective reporting,

and include information guides and clear explanations of correct versus incorrect interpretations of the data when reporting to parents or the general public. Perie, Park, and Klau's (2007) checklist served as clear support for the variables selected for the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study and for the way in which its report handouts were constructed.

2008. *Zwick*. In 2008 there were important findings concerning why educators are not adequately prepared for the data analyses they are expected to conduct. Teachers and administrators must be skilled at using data daily to improve student learning, yet many are not (Zwick et al., 2008). The research team found many teachers and administrators do not know fundamental analysis concepts, and 70% have never taken a college or post graduate course in educational measurement (Zwick et al., 2008).

Few. The statements above were corroborated by Few (2008), a prominent voice in data visualization, who stated that although data is useless if we cannot understand it, most people responsible for analyzing data have received no training to do so. Few also held report design responsible, noting graphs are commonplace today, yet most communicate poorly, and most misinformation graphs communicate is unintentional because charts' creators do not know how to communicate the charts' intended messages (Few, 2008). Few's blog and other writings continue to be highly accessible and informative resources for data system and report vendors.

Underwood. Underwood et al. (2008) were instrumental in voicing concerns over educators' struggles with data and data systems. Teachers do not understand or value some data included in data system reports, and they have difficulty using data systems due to varying technological sophistication levels when it comes to using the data system

to interpret student data, even amongst teachers who serve as assessment coaches to their peers. One of the main problems with using assessment data is that stakeholders at all levels have trouble interpreting the data.

Later, the data system reports were more directly indicated as playing a role in educators' difficulties with data analyses. The reports district administrators are charged with using are presented in ways that are hard for them to read and interpret, and these reports are difficult and time-intensive to analyze (Underwood et al., 2010). Some principals are intimidated by data, and administrators misunderstand the meanings of symbols and terms used in assessment reports and are often confused by the reports' complexity. District administrators often do not have access to data in a format they can use (Coburn, Honig, & Stein, 2009; Underwood et al., 2010).

While PD and staff supports can also help educators' data use, they are not without limitations. Teacher coaches can stop coaching teachers as the school year progresses due to other responsibilities (Underwood et al., 2008). Some assessment coaches are not very tech-savvy and thus have trouble sharing assessments and data system knowledge with teachers. Underwood et al. (2008) found providing a data system designed specifically for users' needs is more effective than expecting training to get users as prepared as they need to be to use the system and its data, and teachers who do not use a data system suggest they would use it on their own if it contained more support for using the data.

A data system can make a huge impact if its design accounts for users of varied data and data system skill levels (Underwood et al., 2008). Underwood et al. (2010) understood the potential, powerful benefits of data, stating data use can lead to insight

into students' abilities and to decisions to improve instruction, and thus offered solutions. Improving a data system's design and reporting can ease some of the growing pains that occur as teachers increase their use of a data system (Underwood et al., 2008). Existing reports can be improved by descriptions to aid understanding of graphics, warnings concerning interpretation limitations, and suggestions for how to apply the data to decision-making (Underwood et al., 2010). Teachers who do not use a data system suggest they would use it on their own if it contained step-by-step instructions as opposed to a complicated help guide, a more user-friendly interface, and information about data available and how this data can be used. While the research team offered numerous, great strides in data system and report design research, they recommended features teachers *feel* will best facilitate their appropriate use and analyses of the data. The important work of Underwood served as support for the variables selected for the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study and paved the way for this study's measurement of the variables' specific impact on data analysis accuracy.

The research team was also instrumental in uncovering how data systems are used in school districts. While some teachers (44%) use the data system directly, others (56%) have access but do not use the data system directly and instead only read printed versions of reports others used the data system to generate (Underwood et al., 2008). Teachers and teacher coaches both report having technology problems such as outdated hardware, inadequate bandwidth, and system freezes, and use of computers outside of the teaching profession influences teachers' success using a data system. For example, most of the teachers in one study reported that they only use the data system to print reports and do not interact with any of the links that accompany the report in the system (Underwood et

al., 2008). Findings concerning ways in which data systems are actually used and ways in which technology can influence data analyses influenced the manner in which respondents' data analyses were facilitated in the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study.

Park. Focusing her case study on high school teachers in urban areas, Park (2008) determined teachers' biases, such as preconceived notions of what they wanted to do with the data and the degree to which they hoped to put these plans into action, impacted the conclusions they drew when making related decisions. Also, their motivation to put the data to use could be hampered by a tendency to blame student performance on factors they perceived to be outside of their control, such as lack of parent involvement, poor student behavior, or lack of resources. Further, pressures of District Office directives, No Child Left Behind requirements, pressure from colleagues, community and parent expectations and viewpoints, and more can also influence the conclusions teachers draw about data they are interpreting and how they put it to use (Park, 2008).

Alverson. In reaction to stakeholders' failure to utilize data to inform some policy and decisions that could possibly benefit from its inclusion, Alverson (2008) explored teacher, administrator, and parent preferences concerning data reports' graphic display types. Alverson was progressive and comprehensive in that she not only explored these stakeholders' preferences, but she also explored what impact the graphic displays had on their ability to accurately garner information from the reports. Unlike the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study, Alverson examined graphic display type rather than the inclusion of data analysis guidance. Alverson's important mixed-method study, which utilized focus groups and a questionnaire, resulted

in recommendations such as simplifying the data when possible, gearing the report toward a specific audience, and defining the task for which the report has been designed.

Alverson's ability to compare preference findings to accuracy findings was also commendable. For example, educators and parents preferred grouped column graphs to segmented/stacked bar graphs, and in this case their preferred graphic display type also rendered more accurate data analyses. Since the report format users report preferring can be the opposite of the reporting format they most accurately interpret (Hattie, 2010), the measuring of accuracy was a vital component to Alverson's (2008) study, and a component many other studies did not address.

2009. *USDEOPEPD.* A national study conducted in relation to No Child Left Behind through the SRI (formerly Stanford Research Institute) International found that teachers are more likely to analyze data by themselves than with their colleagues, and their responses to hypothetical student data suggested they have difficulty with question posing, data comprehension, and data interpretation (USDEOPEPD, 2009). Only 56% of teachers answered correctly in the area of question posing, 64% in data comprehension, and 48% in data interpretation (USDEOPEPD, 2009). These insufficiencies were cited even though the nine school districts studied, involving 18 schools for case studies, were selected for their reputations for strong data use. Thus teachers' struggles witnessed there could be present at other districts. The study was based on the first round of site visits for the national Study of Education Data Systems and Decision Making that ultimately aimed to determine how common education data systems are, how available they are to teachers, their qualities, and their roles in data-driven decisions taking place in schools. Recommendations included providing tools for generating useful data, technical support

for data interpretation, tools for turning data analyses into action, and access to varied levels of data (USDEOPEPD, 2009).

***Knight Commission on the Information Needs of Communities in a Democracy*.** The Knight Commission on the Information Needs of Communities in a Democracy (2009) concluded months of expert deliberations and presentations with recommendations on ways various entities can better assist the public in its acquisition and exchange of information. In its recommendation that communities have online access to pertinent data, the research team stressed the need to not only provide data access, but to also provide an online guidebook that would help users find needed information in the same way a map can help them find physical locales. However, specific recommendations for technology companies focused on discounting products and services rather than making enhancements to better ensure correct data analyses (Knight Commission on the Information Needs of Communities in a Democracy, 2009).

Data literacy refers to one's ability to understand data and ask appropriate questions in relation to it (USDEIESNCEE, 2009). Data literacy also involves viewing data with a critical eye for aspects such as message quality and potential consequences (Knight Commission on the Information Needs of Communities in a Democracy, 2009). However, data literacy must also involve the ability to communicate the data to others and to use the information in some way (Johnson, 2012). Content creation and digital expression play roles in data literacy, and Internet users must use tools that best allow them to process information effectively and draw accurate conclusions from the data they find (Johnson, 2012).

Wayman. This literature review would be grossly incomplete without the inclusion of Wayman (2007), who noted data systems have tremendous potential to assist educators in the inquiry process and help them improve. While much research focuses on PD and staff supports while ignoring the significance of the tool educators are using for data analyses, Wayman drew much attention to the important role data systems play while also acknowledging the gravity of the data analysis error epidemic. It is inadvisable to use data without the assistance of a data system (Cho & Wayman, 2009; Lachat & Smith, 2005). Effective data systems can drastically change how data is used in a school district and can give educators more information in easier ways and in less time than their previous systems (Wayman et al., 2010).

Researchers such as Sanchez, Kline, and Laird (2009) also cited the need for educators to understand how to properly use and analyze their students' data. However, misunderstandings about how to use data and a data system can cripple data use in a school district (Wayman et al., 2009). Teachers might not understand names, labels, and terms used in data systems, and not understanding how to use data can have negative impacts such as low data system use rates and resistance to data. Teachers have frequent difficulties using data, express a need for easier ways to use data, are overwhelmed by data, and have to work longer hours to use data. Wayman also reported many educators struggle to understand how to translate data into specific actions (Cho & Wayman, 2009; Ingram et al., 2004; Supovitz & Klein, 2003; Wayman & Cho, 2009).

Nonetheless, Wayman acknowledged the benefits of common supports. Training and collaboration around the use of educational technology systems can improve data literacy, and educators should turn to PD, better data access, and leadership to improve

educators' ability to turn data into practice (Cho & Wayman, 2009; Knapp et al., 2006; Supovitz & Klein, 2003). When teachers expressed a desire for more support using data, this support included leadership and direction from administrators, training, and support staff to help them understand data; time to collaborate on data can also be helpful (Wayman et al., 2010).

However, Wayman, Cho, and Shaw (2009) also touched on limitations of common data analysis supports, noting most teachers do not collaborate with others when using data, and many teachers do not have enough time to discuss data with others. Meanwhile, other researchers of the time also echoed the limitations of common analysis supports. For example, to help teachers achieve understanding of assessments and their results, learning communities are recommended and go beyond traditional approaches to training (Bennett & Gitomer, 2009). In addition, knowledge management research indicates knowledge can be hard to share with others, even when the intention to share it is there, especially when that knowledge is associated with power or status (Cho & Wayman, 2009). Wayman reported teachers and other educators are quick to take the lead in using data, often operating in front of those planning how they will be supported. Supporting solutions found elsewhere in the research community, the literature noted data systems should be fast and user-friendly (Cho & Wayman, 2009; Wayman et al., 2009; Wayman et al., 2010; Wayman & Stringfield, 2006). Presenting data in more sensible ways is an essential step to improving data use (Cho & Wayman, 2009).

The work of Wayman was monumental in its direct acknowledgement of the responsibility assumed by data system vendors and those involved with them. Data systems' capacity to assist data analysis is unprecedented and the effects on schools are

still being determined (Cho & Wayman, 2009). Data systems can help teachers analyze many sources of data and can help in district-wide initiatives, whereas these and similar endeavors are difficult without such a system. While much of current literature assumes technology serves merely as a tool for decreased hassle and increased productivity, a data system can help with decision-making. Technologies are not simply tools for work; rather, they also influence users' approaches to problem-solving (Cho & Wayman, 2009). However, educators in many districts have difficulty using data, but the issue does not rest on them; rather, it rests on the systems and supports around them, and more needs to be done to help (Wayman et al., 2009).

***Hattie*.** Like his contemporaries, Hattie (2009) noted educators' inadequate data analysis skills, except he highlighted what most proponents of PD and staff solutions ignore: the reports that increasing accountability demands are requiring educators to use are not providing sufficient support to help educators analyze the data they contain. Because there are many readers of reports, any report should include sufficient information to maximize the accuracy of their interpretations. Too much responsibility for making correct interpretations is placed on the test user, whereas more responsibility should be placed on those reporting the test results (Hattie, 2010).

Hattie (2009) was well aware of educators' struggles in interpreting and applying data analyses, noting increasing accountability demands have led to the use of more reports, yet not all users of these reports are as test-sophisticated as they need to be. The U.S.'s No Child Left Behind (NCLB) Act of 2001 led to reports for multiple subjects being distributed at the state, district, school, subgroup, and student levels for parents and teachers of 22 million students per year, yet the reports are not in accordance with any

nationally recognized reporting standards. Most educators are eager to analyze and then act on the data they see, but interpretations require knowledge and understanding (Hattie, 2010; van der Meij, 2008).

Hattie (2009) did concede PD is a viable option. This built on the work of Hattie and Brown (2008), who found those who attend PD more frequently understand score reports and make correct report interpretations that those who do not, and teachers who receive training more accurately comprehend reports than those who do not. However, data systems cannot maximize advantages and minimize detriments unless they provide stakeholders with meaningful feedback. A data system's report design, if done correctly, can free teachers from the need to be assessment literate and instead allow them to focus on instruction and students. For example, teachers better understand assessment results when they are communicated graphically rather than merely numerically. The quality of score reports definitely improves when research is done into how accurately users understand them.

When viewing reports, teachers want more explanations, clearer titles, and more guidance on where to read first (Hattie, 2010). For example, a combination of images and words is more likely to render valid interpretations than numbers, which reduce the likelihood of accurate interpretation when they are overused; thus reports should provide interpretations of numbers. As another example, a shorter, targeted manual or user-friendly Help system causes users to need 40% less training time and to successfully complete 50% more tasks than they would have accomplished with only access to a full-sized manual (Hattie, 2010; van der Meij, 2008). Report validity is dependent on the accuracy and appropriateness with which the report's users interpret and act upon its data,

and validity improves when reports are created or adjusted to increase the accuracy of interpretations. Hattie recommended more research into which reporting principals will increase the accuracy of interpretations being made from them (Hattie, 2010).

Hattie (2009) offered important precautions in applying existing research while highlighting the need for additional research. For example, focus group research, which is the main approach to understanding report interpretations, has shown the report format users report preferring can be the opposite of the reporting format they most accurately interpret. Also, the current trend to include more description and explanation in reports is misleading if the added information is not proven to increase interpretation accuracy. Little research has been done on how users interpret reports (Hattie, 2010).

Hattie (2009) echoed Leeson (2006) in noting that viewing a data system's report on the computer versus printed can negatively impact how it is interpreted; for example, someone who correctly interprets a printed report can make mistakes when scrolling is involved, users are more likely to scan a report on a computer that they would read carefully when printed, and users' inability to mark on the screen can reduce the credibility users attribute to reports. Technology can prevent someone from demonstrating a skill when he or she lacks computer familiarity (Bennett & Gitomer, 2009; Horkay, Bennett, Allen, Kaplan, & Yan, 2006). These findings influenced the manner in which respondents' data analyses were facilitated in the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study.

Hattie has proven to be one of the most respected figures in education research. He eventually consolidated 15 years of research, 50,000 smaller studies, and data on 80

million students to conduct what is known as the world's largest evidence-based study on improving student learning (Hattie & Hargadon, 2013).

Lyrén. Lyrén (2009) also called for more research, noting practitioners and researchers need to gather empirical evidence to support different ways in which subscores are reported in order to prevent misuse. Little or no research has focused on how to best communicate scores on college admission tests to all stakeholders, and there is a need for supporting documentation to assist all users with the advised use and interpretation of scores. Research is needed to address whether students understand their scores and whether results are reported in a way that encourages correct interpretations and prevents misinterpretations (Lyrén, 2009). Allalouf (2009) indicated the failure of some participants to see this need, finding that quality control for test score reporting relates largely to mistakes relating to score validity, such as test examiner or computer errors, and little to do with report design and analysis errors.

2010. ***Zapata-Rivera and VanWinkle.*** Zapata-Rivera and VanWinkle (2010) found teachers need additional help understanding measurement concepts and statistical terms, and adding information to reports can provide this help. In a study involving teachers who had taken at least one course in measurement, all teachers struggled afterwards with statistical terms and measurement concepts and 60% of teachers had difficulty explaining a term used in a score report. The research team also did important work in determining the types of data mistakes teachers were making, and the conditions under which they were making these mistakes. Score reports can more clearly communicate appropriate data-informed actions by including report purpose and use in a way that is easy to comprehend and by adding term definitions, examples, and sample

questions. However, in trying to pinpoint how score reports can more clearly communicate appropriate data-informed actions, the research team interviewed teachers concerning which reports they preferred and recommended adding term definitions, examples, and sample questions to reports (Zapata-Rivera & VanWinkle, 2010). While teacher preference is helpful to know, this research is like that of Underwood et al. (2008) in that its ability to apply theory to practice is limited, as preference is not equivalent to proven effectiveness. Thus the question of how reports can best be improved to enhance analysis accuracy rather than appeal to user preference still remained unanswered.

Data Quality Campaign (DQC). The DQC played an important role in tracking and reporting on data's and data systems' use in the U.S. Launched in 2005 to help states with data systems, the DQC's role involved acknowledgement of the struggles educators experience when trying to use data, noting that few educators automatically know how to use available data effectively (DQC, 2009). The vast majority of educators need guidance in order to understand and use data, including how to apply it to decision-making that can help students succeed. The majority of stakeholders who need to use data to comprehend and raise student achievement are not trained statisticians, and they need additional information to teach them how to use the data they view. Problems persist even at higher levels. For example, states need trained researchers and high-level analysts to make full use of the data they have, yet few states have the resources to add these staff members (DQC, 2009).

The DQC (2010) stressed the benefits of providing data in ways that are easy to interpret correctly and result in better decisions, and it issued guidelines in how growth reports, diagnostic reports, early warning reports, predictive reports, cohort graduation

reports, and feedback reports should be used. In addition, the DQC's three imperative actions to ensure effective data use include verifying data can be analyzed and used, and helping stakeholders apply data to effective decision-making. Reports need to include information like term definitions, how calculations were performed, and data collection details to help users understand report context. Providing educators with access to data is not enough; a data system will not lead to improved student performance unless educators know how to analyze and interpret the data, so they need PD in a variety of formats, including online tutorials on how to use specific reports. Data needs to be provided to users in ways that are easy to interpret and facilitate decision-making. Research and analysis into how to best design reports that are easy to interpret and facilitate decision-making needs to be done by a state or outside agency (DQC, 2010). Other researchers specified related solutions in data systems, such as Kenny (2010), who found that turning information the user selected for analysis into words, accompanied by tables and figures, significantly improved the accuracy of a user's data analysis. While Kenny's work did not address report format or placing analysis guidance into words, it did explore the success of verbalizing analyses, and provided more evidence on how traditional approaches result in improper data analysis despite attempts to teach the user how to analyze data.

The DQC has been consistent in its clear call for action. Providing data will not lead to continuous improvement and student success until practices are in place to help stakeholders throughout the education system understand and properly use the data (DQC, 2009). Policymakers and educators need to consider how day-to-day data use can be better supported. When it comes to data systems at the state level, most attention goes

toward technology aspects like hardware and software, and any reports they generate for educators do not answer questions they could already answer without the system (DQC, 2011). The power of statewide data systems will not be realized until education analysis and researchers – as opposed to just information technology staff – are involved in the full scope of the systems' design and are part of a robust team focused on turning the data into useful information for educators (DQC, 2011).

2011. *National Forum on Education Statistics (NFES).* Other agencies echoed this call to action. For example, the NFES found if data system users do not understand how to properly analyze data, the data will be used incorrectly if it is used at all (NFES, 2011). This statement built on findings that educators sometimes do not know what they need because they are not familiar with what they do not have, such as data system functionalities that can make their jobs easier; stakeholders should ask for a data system that facilitates data analysis and offers extra support to users (NFES, 2010). NFES echoed the popular assertion that PD and in-house data experts are two ways to improve data analysis. However, NFES also noted education leaders should ask what support is available to help staff use data and whether the data analysis tools they are using are user-friendly. For example, to further the impact of PD and in-house data experts, data reports should answer questions, clearly communicate key information, and provide context to guarantee proper interpretation. Since simply *having* data reports and tools is not enough, given educators' propensity for misinterpretation, data systems can include tools to *guide* data use, such as links to instructional materials and guides, in order to help users translate data into instructional actions. Data system resources – such as links to helpful resources, training materials, and video tutorials – can complement traditional training

sessions and guarantee a wide audience's access to training when formal training cannot be provided, ensure new staff members are trained after formal training sessions have passed, and offer training as users' needs evolve (NFES, 2011).

VanWinkle, Vezzu, and Zapata-Rivera. Unlike many applying the theory of how data system supports can improve data analysis accuracy, VanWinkle, Vezzu, and Zapata-Rivera (2011) expanded their focus to administrators and found although administrators are increasingly asked to make data-informed decisions, administrators have trouble understanding data presented in score reports, and score reports designed specifically for administrators are frequently not designed in ways that are easy for administrators to interpret. The research team cited research in which their theory is applied to other non-teacher stakeholders. For example, stakeholders including state politicians, superintendents, and education reporters frequently misunderstand and misinterpret national assessment score reports (Hambleton & Slater, 1996; VanWinkle et al., 2011).

Again, the usual analysis supports were not ignored. PD, leadership, and teacher collaboration should all be used to support effective use of data (VanWinkle et al., 2011; Wayman, 2005). However, the research team appropriately acknowledged additional solutions; for example, recommendations for overcoming stakeholders' unfamiliarity with statistics and statistical terms, as well as their limited time, include field testing reports with targeted audiences and gearing report content and format toward the targeted audiences. Many teachers and administrators use data systems to generate different reports to make decisions, and there is an assumption that data systems require users to already understand the data they are viewing and make the right selections to generate

appropriate score reports. Although some data systems include tools to help users use the system, they usually do not include any support for how to interpret and use the data correctly (Underwood et al., 2010; VanWinkle et al., 2011).

Reports should present information both graphically and textually and include definitions, purpose, use, and cautions concerning interpretation limitations (VanWinkle et al., 2011). Reports should include information that helps users correctly interpret and use the data (Goodman & Hambleton, 2004; Hattie, 2010). Because users' ability to analyze data differs, reports should offer information to cater to users with both beginning and advanced analysis skills, such as through the use of both text and graphics to communicate results (VanWinkle et al., 2011). A link leading to an abstract-like explanation of report components can help users with varied analysis skills better understand the report's terms, interpretations, resources for more information, purpose, and use, as lack of such information can negatively impact the report's use and interpretation. More research is needed on how varied report designs and supports can influence administrators' understanding and use of their contents (VanWinkle et al., 2011).

Odendahl. Odendahl (2011) recognized the importance of minimizing potential misinterpretations, misuses, and negative outcomes otherwise involved in data analysis, and she paid specific attention to the inclusion of supplemental information to help users understand reports, noting that something as complex as test scores cannot be understood without a user's manual. Explanations are needed, such as what the test covers, score meaning, score precision, common misinterpretations, uses for scores, descriptions of skills and knowledge assessed, performance level meanings, peer comparisons,

definitions of essential terms, key data presented in multiple ways, and breakdown of skills and knowledge each student has. Other possible inclusions are sample test questions, sources for more information, suggestions for improving performance, score imprecision aspects. Standards such as AERA can be accommodated between a score report and an interpretation guide. Some data systems feature interactive data reports, but if they linked to more than just variations on data displays they could transform the reports into actual educational tools. More research is definitely needed on test documentation (Moss, 1998; Odendahl, 2011)

Sabbah. In his study on designing more effective accountability communications, Sabbah (2011) used focus groups of parents to examine the best mode of communicating data in school accountability report cards. Educators are starting to realize that data can be the foundation for action toward school improvement, yet few school stakeholders use data to which they have access, and designing public report cards that are easy to interpret is a major challenge. The dependent variable in the study was the degree to which participants comprehended the information and data they were viewing. However, Sabbah's (2011) study contained more independent variables than there would be in a report study involving educators, as the study's parent population was much more varied than educator participants would be, such as in background, education, experience, and language. For example, the parent focus group was composed entirely of native Spanish-speaking parents, which added more complexity and variables.

Sabbah (2011) used the study's results to offer recommendations on how to best improve accountability reports for improved analyses in the public domain. Effective accountability reports must be easy to read and must be accompanied by adequate

interpretive information (Fast & SCASSASR, 2002; Sabbah, 2011). Parents, too, appreciate explanations of what test scores mean and descriptions of skills assessed by tests, and their reports need to contain background, context, recommendations, and clarification. Findings included specific guidelines for accountability reports, tables, bar charts, histograms, and general design, and also reported stakeholder graphic display preferences as including dashboard, bar chart, line graph, stacked line graph with emphasis on gaps, pie chart, and histograms that covered one year rather than three years. These guidelines could benefit the design of accountability report cards for parent consumption, but it is important to note they cannot be automatically applied to other report types or to reports for other users. Fortunately, Sabbah (2011) was aware of these applications and did not overstep the limits of his study when applying the theory to recommended practice; in other words, his recommendations were limited to accountability report cards for parents.

It is important to find a balance between including too much information and too little information on score reports. Like many of his predecessor's, Sabbah (2011) noted the need for more research on the effects of reporting data in education, which is scarce. Educators need to shift from a traditional mindset of gathering report data and focus on steps to better communicate the meaning of assessment report data, such as in ways that invites users to interact with and delve into the information (Sabbah, 2011).

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. Again the U.S. Department of Education produced, through SRI (formerly the Stanford Research Institute), a comprehension study of teachers' data analysis accuracy when using standard student data reports generated with a data system. The

study (USDEOPEPD, 2011) opened by acknowledging teachers are expected to use student data to improve the effectiveness of their practices, which involves the use of student data systems, yet training programs for teachers have generally not addressed data skills and data-informed decision-making. This supported similar assertions by Zwick et al. (2008), Halpin and Cauthen (2011), and others.

Through data analysis questions based on standard data displays, teachers at 13 school districts considered exemplars of active data use, where teachers have access to student data systems and receive support in data-informed decision-making, only achieved 48% correct when making data inferences involving basic statistical concepts such as variability, measurement error or distribution (USDEOPEPD, 2011). Some teachers struggled to make sense of data representations, and a sizeable proportion of teachers made inaccurate inferences when trying to frame data system queries, make sense of differences, or make sense of trends. It was noted it is unlikely teachers at districts where data use is less emphasized would make more accurate data analyses than those described in a study of districts considered exemplars of data use. These findings reflected similar findings of the USDEOPEPD (2009), as well as other studies also covered in this literature review.

2012. *The Bill and Melinda Gates Foundation.* Investigating teachers' use of technology to improve teaching, the Bill and Melinda Gates Foundation (2012) reported on mixed-method opinion research involving focus groups and a national micro-target survey of more than 400 teachers of students in grades six through 12. The Bill and Melinda Gates Foundation found that technological capabilities have not benefitted the U.S. education system – particularly where teachers are concerned – as much as they

have helped U.S. businesses, communication, and lifestyles. Fortunately, teachers indicated overwhelming support for using technology to improve learning, and 85% of teachers reported daily use of technology to support teaching (Bill and Melinda Gates Foundation, 2012).

The Bill and Melinda Gates Foundation cited a need for increases in the usual technology use supports: PD, planning time, peer and coordinator support, and strong leadership at school sites. However, the Bill and Melinda Gates Foundation also indicated ways in which technology companies need to assume increased responsibility in helping their products better contribute to improved learning. Many teachers indicated technology tool companies need to better understand the resource, student, and time challenges teachers face and to do a better job enlisting teacher feedback to make ongoing improvements to their systems. "To be effective, and used to innovate and improve the classroom experience, technology tools must respond to the realities of teacher/student experiences, rather than demand that teachers and students adapt to the requirements of a particular technology" (Bill and Melinda Gates Foundation, 2012, p. 2). Though these statements and the research that contributed to them concerned multiple technologies rather than exclusively data systems, the statements reflect the growing attention that research findings are giving to the need for data systems to do more to ensure educators' appropriate analyses when using data systems and/or their reports to interpret data.

***U.S. Department of Education Office of Educational Technology*.** In an issue brief on using educational data mining and learning analytics to improve teaching and learning, the U.S. Department of Education Office of Educational Technology (USDEOET) (2012) called on educators to ask critical questions about commercial

offerings and purchase intelligently in order to create demand for the most useful educational technology features and uses. This concept is often echoed in less official venues such as blogs, so the publication of this concept by the U.S. Department of Education was an important milestone for improving educational technology products such as data systems. The brief also called for better collaboration between the research, commercial, and educational communities in order to co-design the best educational technology tools, echoing similar messages from the Data Quality Campaign (2011). Also on the note of research, USDEOET (2012) called on researchers and technology developers to conduct research concerning the effectiveness and usability of data displays. This message echoed that of Goodman and Hambleton (2004) Lyrén (2009), Hattie (2010), and others, and mirrored Odendahl's (2011) call for the same type of research in the area of test documentation and data analysis supports. This call for research into the effectiveness and usability of data displays also related to the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study, which investigated varied data displays and the degree to which added analysis supports rendered those displays – and thus the analyses of educators using them – more effective. Discussed within the context of learning systems, USDEOET (2012) also devoted a segment to data visualization resources and functions. The report noted students, parents, teachers, and administrators, who are all data analysis consumers, need data presented in a way that clearly answers questions being posed and points toward a specific and action within the data consumer's means. This acknowledgement of the role of data display related to the premise of this study but was also important as this acknowledgement is rare in field literature, despite the significant influence data display has on educators' data analyses.

The report also made a medicine-to-education analogy, noting labor-intensive data analysis is not a reasonable expectation of educators considering their job demands, so decision support systems need to minimize analysis demands on educators just as such tools do for physicians, as neither profession is more important than the other.

Reporting Standards

Whether or not they are used, national standards have been available over the last two decades to offer guidance concerning the best way to communicate test results (see *Appendix A* for standards applicable to data systems and data system reports). The National Council on Measurement in Education (NCME) issued the Code of Professional Responsibilities in Educational Measurement, which included standards for those interpreting, using, and/or communicating assessment results (National Council on Measurement in Education [NCME], 1995). NCME Standards 6.2-6.5 and 6.8 relay the necessity to accompany reports with additional, non-numeric information to assist with analysis accuracy. For example, Standard 6.2 requires the inclusion of information concerning the assessment on which the data is based, such as its purpose, uses, and limitations, to ensure correct interpretation of the data (NCME, 1995). Standard 6.3 requires a written description of the data that includes proper interpretations and common misinterpretations to avoid. Standard 6.4 requires that results be communicated in a clear way so that intended audiences can understand them; like Standard 6.3, these are also required to include includes appropriate interpretations and likely misinterpretations (NCME, 1995). Standards 6.5 and 6.8 include additional guidelines related to ensuring the appropriate interpretation of results (NCME, 1995). See *Appendix A* for the actual verbiage of each NCME standard.

The American Educational Research Association (AERA) also published standards on how educational testing data should be reported when it issued the Standards for Educational and Psychological Testing (AERA, American Psychological Association [APA], & NCME, 1999). AERA Standards 5.10, 13.1, 13.9, and 13.14 relay the necessity to accompany reports with additional, textual information to assist with analysis accuracy. For example, Standard 5.10 requires that test score information be accompanied by appropriate interpretations that clearly describe, among other things, what the scores mean and common misinterpretations (AERA et al., 1999). Standard 13.1 requires including a clear description of the ways in which the test results should be used, and notes the responsibility of those who require and use tests to recognize and minimize possible problems that could potential arise with their use. (AERA et al., 1999). Standard 13.9 requires that reports used for data-informed decision-making be accompanied by empirical evidence concerning the test scores, instructional programs, and goals for students; if such evidence is not available, the report should include a warning to use multiple measures (AERA et al., 1999). Finally, Standard 13.14 notes that a report should include information on how to interpret the scores, as well as a clear explanation of each score's measurement error (AERA et al., 1999). See *Appendix A* for the actual verbiage of each AERA standard.

The Code of Fair Testing Practices in Education (CFTPE) later presented a Reporting and Interpreting Test Results segment offering guidelines for both test developers and test users (Joint Committee on Testing Practices [JCTP], 2004) (see *Appendix A* for standards applicable to data systems). CFTPE Standards C-1 through C-8 each relate to ways in which reports should include information to help users accurately

interpret results (JCTP, 2004). The list of inclusions is long and includes such details as warnings concerning potential misuse of the data, test benefits and limitations, potential interpretation mistakes, evaluation of test value and performance, and other information that supports recommended interpretations (JCTP, 2004). See *Appendix A* for the actual verbiage of each CFTPE standard.

Despite clear standards on how student data should be reported, educational data systems and their reports do not conform to the standards. Student data reports are not in accordance with any nationally recognized reporting standards (Hattie, 2010). However, the number and scope of these standards add to report content and quantity controversies by recommending an overwhelming number of components. Research notes how easily educators can be overwhelmed by including too much information in reports, so data system vendors need to know which features are most effective in improving analysis accuracy. This controversy and literature concerning it are covered in detail in the *Unanswered Question 1: Content*; *Unanswered Question 2: Quantity*; and *Unanswered Question 3: Impact of Each Component on Analysis Accuracy* sections of this literature review.

Over-the-Counter Labeling Models on Non-Medication Products

Over-the-counter medication's purpose, ingredients, dosage instructions, and dangers are all outlined on a detailed label (Kuehn, 2009). This allows patients to take over-the-counter medication with the goal of improving wellbeing while a doctor is not present. No or poor medication labels have resulted in many errors and tragedy, as people are left with no way to know how to use the contents wisely (Brown-Brumfield & DeLeon, 2010). Similarly, many data systems display data for educators without

sufficient support to use their contents – data – wisely (Coburn, Honig, & Stein, 2009; Data Quality Campaign [DQC], 2009, 2011; Goodman & Hambleton, 2004; National Forum on Education Statistics [NFES], 2011).

Fortunately, research indicates label conventions can result in improved understanding on non-medication products, as well (Hampton, 2007; Qin et al., 2011). Like Kuehn (2009) and DeWalt (2010), Hampton (2007) reported on the topic of labeling in *The Journal of the American Medical Association*. Arguing that the absence of labeling makes it difficult for users to determine which devices are safe for particular patients, the American Medical Association, American Nurses Association, and other health care organizations urged the FDA to require mandatory labeling on medical devices containing chemicals found to be harmful to particular populations (Hampton, 2007). Even though the devices were deemed safe for some populations, some hospitals did away with the non-labeled devices entirely (Hampton, 2007).

Qin et al. (2011) interviewed 876 adults concerning the impact of cigarette warning labels on their understanding of smoking dangers, likelihood of giving cigarettes to others, and likelihood of quitting smoking. While Chinese labeling was found to be insufficient, foreign label formats such as those used in Canada had a positive impact in conveying information such as health warnings (Qin et al., 2011). Qin et al. (2011) found warning labels combining text with graphics were more effective than text-only labels, clear and direct messages worked best, countries are increasingly mandating better graphic imagery and warning labels on cigarette packaging, and labels with detailed risk information and graphics were more effective in deterring unhealthy behavior. Approximately 33% of smokers reported they were likely to quit smoking due to the

warning labels with graphic and detailed cessation and health risk information (Qin et al., 2011).

If passed, the HR 3553 Bill would amend the Federal Food, Drug, and Cosmetic Act, Federal Meat Inspection Act, and Poultry Products Inspection Act to add the requirement that food containing or produced with genetically engineered (GE) material be labeled correspondingly (Open Congress, 2011). Clay (2012) reviewed the status of the Genetically Engineered Food Right to Know Act (HR 3553), noting the bill would also require the FDA to test products periodically for compliance with the labeling legislation, as well as institute a framework to ensure the accuracy of labeling. In making a case for the non-medication labeling bill, Clay (2012) noted more than 90% of surveyed Americans have expressed support for the labeling of GE foods, and this rate of support has been maintained for 20 years.

Though for products other than medication, Hampton (2007), Qin et al. (2011), and Clay (2012) offered or called for label recommendations similar to those recommended by the FDA for over-the-counter medication labels. The FDA directs the pharmaceutical industry to accompany nonprescription medications with clear and accurate instructions, which should be tested for usability to ensure that patients of all literacy levels can accurately use them. This leads to greater customer safety and satisfaction, the cost to test these precautions' effectiveness is miniscule, and until such effectiveness is tested a product is relying on face value and opinion rather than solid evidence on how effective its labeling is in reducing errors (DeWalt, 2010). DeWalt, (2010) noted drug companies are responsible for making information like dose indications clear to patients. A physician's work in diagnosing a patient and prescribing

treatment is meaningless if the patient cannot use the medication properly by the time he or she obtains it; assuming the patient will understand how to use the medication is a mindset that has been outdated for decades (DeWalt, 2010). Also, research into the guidance's effectiveness is paramount. Providing guidance is a good starting point, but proceeding without evidence that the guidance eliminates errors is negligent (DeWalt, 2010). DeWalt's (2010) premise comprises part of the motivation behind the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study.

Likewise, Hampton's (2007), Qin et al.'s (2011), and Clay's (2012) non-medication label recommendations were similar to those recommended by Watanabe, Gilbreath, and Sakamoto (1994) for over-the-counter medication labels. A Senior Assembly Proposal presented to the California Assembly, based on a the recommendations of a panel of optometrists and ophthalmologists involved in a New England College of Optometry study, called for making over-the-counter medication labels more legible though size of at least 1.2mm in vertical height and no more than 40 characters per inch, noting other legibility factors than type size include letter contrast, line spacing, print and background color, and type style (Watanabe, Gilbreath, & Sakamoto, 1994). These recommendations were made to improve administration accuracy when the contents of over-the-counter medication packaging are consumed. This *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study is based on the premise that just as over-the-counter medicine's proper use is communicated with a thorough label, and just as over-the-counter label conventions can be applied to non-medication products, a data system used to analyze student performance can include components that might help users better comprehend the data it contains.

Behavioral Economics and Data-Informed Decision-Making

Simon (1979) purported that people are *boundedly rational* when making decisions. The concept of bounded rationality was introduced as an alternative to rational analysis and proposed that people behave in ways that are nearly optimal, as opposed to completely optimal, as their resources will allow when seeking their goals (Simon, 1979). Behavioral Economics grew from an attempt to map bounded rationality by investigating how systematic biases result in differences between people's decisions and the optimal decisions made in rational-agent models (Kahneman, 2003). Simply put, behavioral economics accounts for the fact that people do not always behave rationally.

Hundreds of studies verify that people's decision-making is inherently biased and otherwise flawed (Thaler & Sunstein, 2008). For example, Park (2008) found teachers' biases, such as preconceived notions of what they wanted to do with the data and the degree to which they hoped to put these plans into action, impacted the conclusions they drew when making data-informed decisions. Factors such as emotions, instincts, biases, and loss aversion can all cause people to make decisions that are less than completely rational (Kahneman, 2011; Lehrer, 2011). More research is needed on how an adjustment to one of these constructs influences the impact of the others (Mitchell, 2010). Nonetheless, the format through which context is organized further influences decision-making, and companies should take advantage of opportunities to influence this decision-making in ways that will improve people's lives (Thaler & Sunstein, 2008). When companies are those that provide student data systems, the lives most significantly impacted by such improvement are students, though other stakeholders can also benefit.

Though the full extent of their import was largely unrecognized at the time, the writings of Thaler (1980) and of Kahneman, Slovic, and Tversky (1982) laid some of the earliest groundwork in behavioral economics. For example, it is now largely accepted that the organization of the context within which people make decisions – termed *choice architecture* –impacts decision-making (Thaler & Sunstein, 2008). However, it was not until the early 1990s that behavioral economics received widespread acceptance as a field somewhere between psychology with economics (Camerer, Loewenstein, & Rabin, 2003). Now the application of behavioral economics is widespread. For example, Horizon Blue Cross Blue Shield of New Jersey is targeting behavioral economics in attempts to engage consumers in healthcare delivery (Wood, 2012), and the discipline is being used to justify government intervention to combat obesity (Marlow & Abdukadirov, 2012).

Behavioral economics' import on education is especially noticeable where the use of student data is concerned, as educators analyze student data in conjunction with decision-making processes. In the 2000s, educators of all levels increasingly embraced data-informed decision-making, which involves systematically analyzing data to guide decisions aimed at helping students succeed (Marsh et al., 2006). Behavioral economics involves two systems of thinking and decision-making (Kahneman, 2003, 2011). These two systems can be classified as the *Automatic System* (System 1), which is intuitive, and the *Reflective System* (System 2), which is *rational* (Thaler & Sunstein, 2008). Both systems control a person's attention, and one system must borrow attention from the other when required since a person's attention capacity is limited, such as enlisting more System 2 thought processes when the brain is in analytic mode and undergoing cognitive

strain (Kahneman, 2011). The analysis of student data is a clear example requiring System 2, particularly for educators who find the analysis to be unfamiliar or difficult. While each of the two systems contributes to decision-making, the process of thinking and deciding is also influenced by factors such as priming, biases, heuristics, prototypes, judgments, anchoring, and framing (Kahneman, 2011). For example, small and seemingly insignificant differences in how content is arranged can mean a significant difference in the decisions people make based on that content (Thaler & Sunstein, 2008).

Priming. Priming is a dimension of behavioral economics that involves one idea resulting in another, among many. Basically, a subtle influence such as a hint of an idea primes one's thoughts, which then impact one's actions in ways that can be surprisingly significant (Thaler & Sunstein, 2008). Virtually anything can serve as a priming source, such as a word, an action, or a gesture (Kahneman, 2011). Behavioral economics research on priming is relegated to conclusions concerning tendencies, and not enough is known about how priming is influenced by real-life factors such as people changing and learning (Mitchell, 2010).

When applied to data-informed decision-making, an important source of priming can involve resources educators interact with before viewing data to inform their decisions: the resources prime the educator's thoughts concerning the data, and then those thoughts prime the educator's decisions. For example, Goodman, and Hambleton (2004) noted the value gained in states that accompanied data reports with information for parents to read *before* reading and interpreting the data. More subconscious priming sources involved in data-informed decision-making include the environment in which

analyses take place and the individual's associations with those facilitating the session, the data system used and the individual's associations with technology, etc.

Biases, heuristics, prototypes, and judgments. Research confirms that decisions people make are inherently flawed due to factors such as bias (Thaler & Sunstein, 2008). For example, the institutional nature of military decision-making processes (MDMP), organizational culture, and individuality all impact the heuristics and biases that influence how military commanders respond to surprises while in action (Williams, 2010). System 1 thought processes use biases and heuristics, such as prototypes, to speed up thinking and decision-making; a social example of this is a stereotype, which does not necessarily lead to an accurate conclusion (Kahneman, 2011).

Biases, heuristics, and prototypes, as well as the judgments to which they lead, are not always undesirable. For example, if someone is walking down a dark alley and a van with tinted windows pulls up, it would be wise to avoid the van. However, biases, heuristics, and prototypes can also cause flawed judgments, such as where data-informed decision-making is concerned. For example, urban high school teachers' biases in the form of preconceived notions impacted the conclusions they drew when making data-informed decisions (Park, 2008).

Anchoring. An anchor is a value someone considers before estimating the quantity of something, such as a home's asking price, and anchoring effect is the phenomenon that causes his or her estimate to stay closer to the anchor than it might have been is the anchor were not considered (Kahneman, 2011). Anchoring usually results in an estimate that does not match reality (Williams, 2010). Thus anchoring is another example of a heuristic (Thaler & Sunstein, 2008). Anchoring can occur in data-informed

decision-making when educators have preconceived notions of an entity's performance. For example, if Teacher A heard in the staff room that 68% of her school's students passed the state graduation test, her analysis of state graduation test data might later be skewed – or *anchored* – by her consideration of the statistic she heard in the staff room. Essentially, the anchor primed the teacher's thoughts, which then primed her actions. Thus the skewed analysis Teacher A made is likely to result in skewed data-informed decision-making.

Framing. The manner in which content is organized for people using it to make decisions significantly impacts those decisions (Thaler & Sunstein, 2008). Framing applies to how information is presented, as presenting the same information to someone in different ways will often result in different emotions and different levels of difficulty in understanding or analyzing the information (Kahneman, 2003, 2011). Framing thus plays a large role in data analysis accuracy and data-informed decision-making (see the *Chapter 2: Literature Review: History of Specific Research Contributions* section for a historic timeline of research-based recommendations for report design, which relate to framing). Thus the reports used in this *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study subscribed to leading research-based recommendations concerning the best ways in which to present the data in report format, though they did so in a way that did not deviate from what is commonly seen in data systems currently on the market. In other words, reports used in the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study adhered to the better data presentations commonly seen in data systems, but they did not adhere to the best data presentations that – despite being more effective – are not yet commonly seen in student data systems.

Suggested ways to present analysis guidance in footers, abstracts, and interpretation guides were also utilized in the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study, but the best manner in which to frame these resources had not yet been determined in regards to direct impact on analysis accuracy. Thus each of the three support resources were framed in two different formats for respondents in the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study.

The Current State of Educators' Data Analysis Skills

Educators must be skilled at using data daily to improve student learning, yet many are not (Zwick et al., 2008). Misunderstandings about how to use data and a data system can cripple data use in a school district and cause low data system use rates and resistance to data (Wayman et al., 2009). Unfortunately, not all educators have the skills needed to successfully use data to inform decisions, and having data does not mean it will be used properly (Marsh et al., 2006). Few educators automatically know how to use available data effectively (DQC, 2009).

For example, educators' incomplete understanding of statistics can lead them to draw false conclusions from data (Marsh et al., 2006). Many teachers and administrators do not know fundamental analysis concepts, and 70% have never taken a college or post graduate course in educational measurement (Zwick et al., 2008). Few teacher preparation programs cover topics like state data literacy (Halpin & Cauthen, 2011; Stiggins, 2002). Training programs for teachers have generally not addressed data skills and data-informed decision-making (USDEOPEPD, 2011). In fact, most people

responsible for analyzing data have received no training to do so (DQC, 2009; Few, 2008).

Many educators experience difficulties just trying to understand the data they are analyzing (Goodman & Hambleton, 2004; Hambleton, 2002; Hattie, 2010; NRC, 2001). Teachers have frequent difficulties using data, express a need for easier ways to use data, and are overwhelmed by data, (Wayman et al., 2010). For example, teachers have difficulty using data systems due to varying technological sophistication levels when it comes to using the data system to interpret student data, even amongst teachers who serve as assessment coaches to their peers (Underwood et al., 2008). The problem is not restricted to teachers. Stakeholders at all levels have trouble interpreting data, such as principals who are intimidated by data and need training, and teacher coaches who are not tech-savvy and have trouble sharing assessments and data system knowledge with teachers (Underwood et al., 2008). State-level stakeholders are also at varying stages of being able to actually analyze the data that data systems display (Minnici & Hill, 2007). Even at the state level, stakeholders are not using student data effectively (Halpin & Cauthen, 2011). However, if data system users do not understand how to properly analyze data, the data will be used incorrectly if it is used at all (NFES, 2011).

One of the most comprehensive studies on the topic of teacher data analysis accuracy, which was conducted for the U.S. Department of Education Office of Planning, Evaluation and Policy Development (USDEOPEPD) (2009) in relation to NCLB, involved case studies of 18 schools in nine school districts that were selected for their reputations for *strong* data use. Despite these promising reputations, researchers found teachers' responses to hypothetical student data suggested they have difficulty with

question posing, data comprehension, and data interpretation (USDEOPEPD, 2009). Teachers answered 44% of questions incorrectly in the area of question posing, 36% incorrectly in data comprehension, and 52% incorrectly in data interpretation (USDEOPEPD, 2009). The study was based on the first round of site visits for the national Study of Education Data Systems and Decision Making that ultimately aimed to determine how common education data systems are, how available they are to teachers, their qualities, and their roles in data-driven decisions taking place in schools (USDEOPEPD, 2009).

Likewise, a study of teachers at 13 school districts considered exemplars of active data use, where teachers have access to student data systems and receive support in data-informed decision-making, rendered scores of 48% correct when making data inferences involving basic statistical concepts such as variability, measurement error or distribution (USDEOPEPD, 2011). Some teachers struggled to make sense of data representations, and a sizeable proportion of teachers made inaccurate inferences when trying to frame data system queries, make sense of differences, or make sense of trends. Given that these insufficiencies were found at districts known for *strong* data use, teachers' struggles witnessed there are likely present at other districts. It is unlikely teachers at districts where data use is less emphasized would make more accurate data analyses than those described in a study of districts considered exemplars of data use (USDEOPEPD, 2011).

Controversy Concerning the Best Way to Improve Data Analysis Accuracy

Most educators are eager to analyze and then act on the data they see, but they cannot correctly interpret it when they do not have the required knowledge and understanding to do so (van der Meij, 2008). Many theories have surfaced on how to

provide educators with the knowledge and understanding needed to improve the accuracy of their data-informed conclusions. Two of these theories dominated most literature on the topic. One theory is PD can improve educators' data analysis accuracy (Lukin et al., 2004; Sanchez et al., 2009; Zwick et al., 2008). The other prevailing theory is staff – such as site leaders, data teams, data experts, and/or instructional coaches – can improve educators' data analysis accuracy (Bennett & Gitomer, 2009; McLaughlin & Talbert, 2006). Most experts supported both of these two theories, recommending both PD and staffing to improve data use (Marsh et al., 2006; NFES, 2011; USDEOPEPD, 2009; and VanWinkle et al., 2011). Receiving less but growing attention is a third theory: supports *within* data systems can improve educators' data analysis accuracy (Hattie, 2010; Underwood et al., 2010; Wayman et al., 2010; Zapata-Rivera & VanWinkle, 2010).

Supports Outside of Data Systems Are Not Enough

As an example of dominating research themes, recommendations by authorities such as the U.S. Department of Education for using data to support instructional decisions focused on PD, accessing data from multiple sources, site-based data teams, and data discussions, and overlooked the prospect of including analysis guidance within the data system, adding tools to improve data interpretations are missing from most data systems (USDEOPEPD, 2009). The call for making data systems and their reports share the responsibility of improving educators' data interpretations signaled a monumental shift in data skills research (Hattie, 2010). Historically, the burden of boosting educators' data skills was placed on PD and staff resources. While experts concluded those two approaches are beneficial, PD and staff resources are not enough. Data systems do not include proper support for interpreting data and turning results into action, despite the

fact that teachers do not often know how to translate data into action (Rennie Center for Education Research and Policy [RCERP], 2006). This could become the biggest challenge facing effective data use once educators are accessing technology otherwise deemed adequate (RCERP, 2006).

For example, those who receive PD can continue to struggle. The most common method of supporting data-informed decision-making is PD focused on understanding test data, but its value varies, the majority of teachers and principals do not find it to be helpful, and sessions do not typically cover how to use test results for instructional planning (Marsh et al., 2006). In one study involving teachers who had taken at least one course in measurement, *all* teachers struggled afterwards with statistical terms and measurement concepts (Zapata-Rivera & VanWinkle, 2010).

Likewise, staff supports do not always operate as intended. Knowledge management research indicated knowledge can be hard to share with others, even when the intention to share it is there, especially when that knowledge is associated with power or status (Cho & Wayman, 2009). Site leaders are another source of data analysis support, but the quality of leadership varies (Marsh et al., 2006). Also, teachers and other educators are quick to take the lead in using data, but in doing so they often operate in front of those planning how they will be supported (Wayman et al., 2010; Wayman & Stringfield, 2006). Problems persist even at higher levels; states need trained researchers and high-level analysts to make full use of the data they have, yet few states have the resources to add these staff members (DQC, 2009).

In addition, district budgets are limited, and while selecting a data system with analysis support over one without does not have to cost added funds, PD and additional

staffing typically do. While teachers feel more comfortable with in-person PD, this training format is expensive and research showed that single-day workshops do not significantly alter teacher behavior (Fletcher, 2012). Translating data into action is complex, and in order to effectively use data analysis tools teachers will need ongoing support; these are offered in the form of coaches and PD, but at a cost (Rennie Center for Education Research and Policy, 2006). Data staff and training resources can be limited at the local level, as is staff with proper data analysis experience and skills at the state level (McDonald et al., 2007).

Even when budgets do allow for extensive PD and support staff, these supports are not ever-present. Most teachers are making instructional decisions based on data they view while alone (USDEOPEPD, 2009), helping to explain why even staff at a district with the funds for PD and added staffing continues to draw incorrect "data-informed" conclusions. Most teachers do not collaborate with others when using data, and many teachers do not have enough time to discuss data with others (Wayman et al., 2009). Thus a data system can serve as a virtual data coach when colleagues or trainers are not present. Providing a data system that is designed specifically for its users' needs is more effective than expecting training to get users as prepared as they need to be to use the system and its data (Underwood et al., 2008). There is a clear need for research identifying how assessment results can most effectively be reported (Goodman & Hambleton, 2004; Hattie, 2010). However, research promoting analysis supports within data systems left some questions unanswered.

Unanswered Question 1: Content

Experts agreed more could be asked of data systems to improve data analyses (DQC, 2010; Lyrén, 2009; Odendahl, 2011; Wayman et al., 2010). However, there were conflicting viewpoints as to how the data system can best do this. One unanswered question relates to *content*. When adding data analysis guidance to a data system, providers need to know what supporting text should contain. The recommendations were vast, yet experts also cautioned against including everything. Consider the following sampling of recommendations:

Underwood, Zapata-Rivera, and VanWinkle (2010) suggested enhancing existing reports with descriptions to aid understanding of graphics, warnings concerning interpretation limitations, and suggestions for how to apply the data to decision-making. DQC (2009) stated reports need textual information like how calculations were performed and data collection details to help users understand report context. VanWinkle, Vezzu, and Zapata-Rivera (2011) suggested including purpose, use, and cautions concerning interpretation limitations. Term definitions should also be included (Wayman et al., 2009; Zapata-Rivera & VanWinkle, 2010). Others called for explanations, such as what the test covers, score and performance level meanings, descriptions of skills and knowledge assessed, score precision, common misinterpretations, uses for scores, and a breakdown of the skills and knowledge each student has (Odendahl, 2011). For example, accountability reports should contain adequate interpretive information, including cautions concerning possible misinterpretations, and should be designed with the goal that even one's next-door-neighbor should understand their meaning (Fast & State Collaborative on Assessment and Student Standards Accountability Systems and

Reporting Consortium [SCASSASRC], 2002). Zapata-Rivera and VanWinkle (2010) recommended including report purpose and use.

If text added to reports could accommodate all the above recommendations, there would be no controversy. However, the current trend to include more description and explanation in reports is misleading if the added information is not proven to increase interpretation accuracy, and though not enough research has been done in this area, reports should rely more on visuals than text (Hattie, 2010). Too much information or text can overwhelm users and cause them to miss higher-level implications (Hattie, 2010; VanWinkle et al., 2011; Zapata-Rivera & VanWinkle, 2010). Despite recommendations for added text, effective score reports are clear, concise, easy to read, and *jargon*-free (Odendahl, 2011). In addition, some stakeholders do not like reports that are too technical and contain complex definitions, and it is important to find a balance between including too much information and too little information on score reports (Sabbah, 2011). Even under ideal circumstances, social science research confirms that people do not always make the best use of resources, even when that use directly impacts their wellbeing (Thaler & Sunstein, 2008). Thus practitioners must balance the many research-based requests for added text and strive to include only the text that future research deems most pertinent.

Unanswered Question 2: Quantity

Data system providers also need to know what the right *quantity* of guidance is in order to help but not overwhelm the user, as too much added information can overwhelm the audience, rendering the guidance unused and thus useless. As seen above, many studies had long lists of items that systems and reports should contain, but equally

prevalent was research noting how easily educators can be overwhelmed by such information. Consider the following recommendations:

VanWinkle et al., (2011) considered differing users' abilities to analyze data when suggesting reports should offer information to cater to users with both beginning and advanced analysis skills, such as through the use of both text and graphics to communicate results, and should provide guidance in how to make the right selections to generate appropriate score reports. Others suggested a glossary of terms and other interpretive information (Goodman & Hambleton, 2004; Hattie, 2010). Help manuals and guides were also popular. Experts recommended offering a short, targeted manual (Hattie, 2010; van der Meij, 2008). However, teachers who do not use a data system suggested they would use it on their own if it contained step-by-step instructions as opposed to a complicated help guide, a more user-friendly interface, and information about data available and how this data can be used (Underwood et al., 2008).

Others suggested interpretation guides. Many experts agreed systems should include support for how to interpret and use the data correctly (Lyrén, 2009; Odendahl, 2011; Underwood et al., 2010; VanWinkle et al., 2011). The Council of Chief State School Officers Accountability Systems and Reporting State Collaborative checklist for communicating accountability results recommended reports communicate all relevant data clearly, promote accurate interpretation and use of data, use a format that helps schools learn how to use the data, apply the latest research in effective reporting, and include information guides and clear explanations of correct versus incorrect interpretations of the data when reporting to parents or the general public (Perie et al., 2007).

Experts also recommended offering sample test questions, sources for more information, suggestions for improving performance, score imprecision aspects (Goodman & Hambleton, 2004; Odendahl, 2011). Zapata-Rivera and VanWinkle (2010) also recommended including examples and sample questions. National Forum on Education Statistics (2006) noted to truly be a decision support system, a data system needs robust reporting tools that can include explanatory information within charts, legends, citations, explanations, and other information to clarify the data's meaning.

Links are also popular. Experts noted a link leading to an abstract-like explanation of report components can help users with varied analysis skills better understand the report's terms, interpretations, resources for more information, purpose, and use, as lack of such information can negatively impact the report's use and interpretation (Goodman & Hambleton, 2004; NFES, 2011; VanWinkle et al., 2011). Data system links to helpful resources, training materials, and video tutorials can complement traditional training sessions and guarantee a wide audience's access to training when formal training cannot be provided, ensure new staff members are trained after formal training sessions have passed, and offer training as users' needs evolve (NFES, 2011). The Data Quality Campaign (2009) also stated a data system will not lead to improved student performance unless educators know how to analyze the data, so online tutorials on how to use specific reports are needed.

Once again, practitioners must balance the many research-based requests for added features and strive to include only features that future research deems most pertinent. Data systems and their reports should include whatever information that helps users correctly interpret and use the data (Fast & SCASSASRC, 2002; NFES, 2011;

Sabbah, 2011; Tufte, 2011). However, to help practitioners choose between research-based recommendations for data system text and features, more research is needed on how these variables can help in the data's analysis (Goodman & Hambleton, 2004).

Unanswered Question 3: Impact of Each Component on Analysis Accuracy

Since not every recommendation may be accommodated (as other recommendations discourage including too many supports), research must determine how likely each data system support is to increase analysis accuracy – essentially, how various recommendations compare to one another in effectiveness. Historically, some researchers sought to address this issue. Aschbacher and Herman (1991) offered some help with this controversy, suggesting that a report be balanced and devote space to explanations based on their importance. However, the questions of which components are most important and where the cutoff for space occurs remained unanswered. Unfortunately, construct validity is often weak in studies of report format in education, as studies often involve case studies or focus groups that are used to examine which reports educators *prefer* or which reports educators *identify* as helpful. This means researchers are examining participant preference and opinion but not necessarily report success rates.

For example, Underwood et al. (2008) noted teachers do not understand or value some data included in data system reports and have difficulty using data systems due to varying technological sophistication levels, even amongst teachers who serve as assessment coaches to their peers. Underwood et al. (2008) found providing a data system designed specifically for users' needs is more effective than expecting training to get users as prepared as they need to be to use the system and its data, and teachers who do not use a data system suggest they would use it on their own if it contained more

support for using the data. Underwood et al. (2008) recommended features teachers *feel* will best facilitate their appropriate use and analyses of the data. However, while teacher preference and opinion is helpful to know, other research noted the approach's limitations when it comes to applying the results to practice. Focus group research, which is the main approach to understanding report interpretations, showed report format users report preferring could be the opposite of the reporting format they most accurately interpret (Hattie, 2010). Thus the question of how reports can best be improved to enhance analysis accuracy rather than appeal to user preference remained unanswered.

As another example, Zapata-Rivera and VanWinkle (2010) found teachers need additional help understanding measurement concepts and statistical terms, and adding information to reports can provide this help. In a study involving teachers who had taken at least one course in measurement, all teachers struggled afterwards with statistical terms and measurement concepts and 60% of teachers had difficulty explaining a term used in a score report (Zapata-Rivera & VanWinkle, 2010). Zapata-Rivera and VanWinkle (2010) did important work in determining the types of data mistakes teachers were making, and the conditions under which they were making these mistakes. However, in trying to pinpoint how score reports can more clearly communicate appropriate data-informed actions, Zapata-Rivera and VanWinkle (2010) interviewed teachers concerning which reports they *preferred* and recommended adding term definitions, examples, and sample questions to reports (Zapata-Rivera & VanWinkle, 2010). While teacher preference is helpful to know, this research is like that of Underwood et al. (2008) in that its ability to apply theory to practice is limited, as preference is not equivalent to proven effectiveness.

Thus the question of how reports can best be improved to enhance analysis accuracy rather than appeal to user preference still remained unanswered.

Summary

Educators' data-informed decisions can improve student learning (Sabbah, 2011; Underwood et al., 2010; Wohlstetter et al., 2008). Research reveals that most educators have access to data systems to generate and analyze student score reports (Aarons, 2009; Herbert, 2011), and educators use data systems to make decisions that impact students (VanWinkle et al., 2011). However, literature also features evidence educators do not use this data correctly, and there is clear evidence many users of data system reports have trouble understanding the data (Hattie, 2010; NRC, 2001; Wayman et al., 2010; Zwick et al., 2008). For example, in a national study of districts known for *strong* data use, teachers incorrectly interpreted data in 52% of instances (USDEOPEPD, 2009). It is unlikely teachers at districts where data use is less emphasized would make more accurate data analyses than those described in a study of districts considered exemplars of data use (USDEOPEPD, 2011).

Possible causes for these inadequacies include the facts that few teacher preparation programs cover topics like assessment data literacy (Halpin & Cauthen, 2011; Stiggins, 2002), and most people responsible for analyzing data received *no* training to do so (DQC, 2009; Few, 2008). While literature supports PD and staff supports as potential sources of improved data analysis accuracy, literature also indicates these approaches are not exhaustive, as both have limitations. For example, PD's value varies, the majority of teachers and principals do not find it to be helpful, and sessions do not typically cover how to use test results for instructional planning (Marsh et al., 2006). In one study

involving teachers who had taken at least one course in measurement, all teachers struggled afterwards with statistical terms and measurement concepts (Zapata-Rivera & VanWinkle, 2010). In-person PD is expensive and research showed that single-day workshops do not significantly alter teacher behavior (Fletcher, 2012). Data staff and training resources can be limited at the local level, as is staff with proper data analysis experience and skills at the state level (McDonald et al., 2007). Even when budgets do allow for extensive PD and support staff, these supports are not ever-present. Most teachers do not collaborate with others when using data, and many teachers do not have enough time to discuss data with others (Wayman et al., 2009). Most teachers are making instructional decisions based on data they view while alone (USDEOPEPD, 2009), helping to explain why even staff at a district with the funds for PD and added staffing continues to draw inaccurate "data-informed" conclusions. Research contains evidence providing a data system that is designed specifically for its users' needs is more effective than expecting training to get users as prepared as they need to be to use the system and its data (Underwood et al., 2008). The process of thinking and deciding is also influenced by factors such as priming, biases, heuristics, prototypes, judgments, anchoring, and framing (Kahneman, 2011). Even small and seemingly insignificant differences in how content is arranged can mean a major impact on the decisions people make based on that content (Thaler & Sunstein, 2008). This means data-informed decision-making is influenced by these behavioral economics dimensions, so added data analysis supports embedded within a data system must also have their optimal framing formats determined.

Growing research attention is devoted to data systems' role in the data analysis process. Data use impacts students, and misunderstandings when using data systems can

cripple data use (Wayman et al., 2009). Literature indicates data systems do not include proper support for interpreting data and turning results into action (RCERP, 2006). Many data systems display data for educators without sufficient support to use their contents – data – wisely (Coburn et al., 2009; DQC, 2009, 2011; Goodman & Hambleton, 2004; NFES, 2011).

Meanwhile, research indicates many benefits of over-the-counter medication labeling. For example, missing or inadequate medication labels have resulted in many errors and tragedy, as they leave people with no way to know how to use the contents wisely (Brown-Brumfield & DeLeon, 2010). Hampton (2007), Qin et al. (2011), and Clay (2012) offered or called for label recommendations similar to those recommended by the FDA for over-the-counter medication labels. Research contains evidence label conventions can result in improved understanding on non-medication products, as well (Hampton, 2007; Qin et al., 2011). Despite this, labeling and tools within data systems to assist analysis are uncommon, even though most educators analyze data alone (USDEOPEPD, 2009).

Thus the prospects of applying over-the-counter medication labeling benefits to data systems is worthy of exploration. However, research promoting analysis supports within data systems left key questions unanswered. For example: (a) there are conflicting findings concerning what additional analysis information should be included with data reports, leaving the question of *content* unanswered; (b) research contains evidence not all recommendations should be included on or with reports since the magnitude can overwhelm educators and lead to less success than the inclusion of fewer details, leaving the question of *quantity* unanswered; and (c) since not every recommendation may be

accommodated, and research is needed to determine how likely each data system support is to increase analysis accuracy, the question of each component's *impact* has been left unanswered. Literature notes a clear need for research specifically identifying how reports can better facilitate correct interpretations by its users (Goodman & Hambleton, 2004; Hattie, 2010). The full potential of data systems that generate these reports will not be reached until researchers contribute to improving data system design to improve data analyses (DQC, 2011).

Chapter 3: Research Method

The problem investigated was educators make data analysis errors impacting students, yet data systems and reports do not include analysis help, and it was undecided whether adding supports to data systems can reduce the number of analysis errors. Data-informed decisions can improve learning (Sabbah, 2011; Underwood, Zapata-Rivera, & VanWinkle, 2010; Wohlstetter, Datnow, & Park, 2008). Educators worldwide test students, distribute score reports, and expect stakeholders to make improvements based on these reports (Hattie & Brown, 2008). Most educators have access to data systems to generate and analyze score reports (Aarons, 2009; Herbert, 2011).

Unfortunately, educators do not use this data correctly, and there is clear evidence many users of data system reports have trouble understanding the data (Hattie, 2010; National Research Council, 2001; Wayman et al., 2010; Zwick et al., 2008). For example, in a national study of districts known for *strong* data use, teachers incorrectly interpreted 52% of data (USDEOPEPD, 2009). Few teacher preparation programs cover topics like assessment data literacy (Halpin & Cauthen, 2011; Stiggins, 2002), most people analyzing data received *no* training to do so (DQC, 2009; Few, 2008), and human biases compromise judgment and complicate decision-making processes (Kahneman, 2011).

Data use impacts students, and misunderstandings when using data systems can cripple data use in school districts (Wayman, Cho, & Shaw, 2009). Yet labeling and tools within data systems to assist analysis are uncommon, even though most educators analyze data alone (USDEOPEPD, 2009). There is a clear need for research identifying how reports can better facilitate correct interpretations by its users (Goodman & Hambleton, 2004; Hattie, 2010). The power of data systems that generate these reports

will not be realized until researchers contribute to improving data system design to improve analysis (DQC, 2011).

The purpose of this experimental quantitative study, conducted in a laboratory environment, was to facilitate causal inferences concerning the degree to which including different forms of data usage guidance within a data system reporting environment can improve educators' understanding of the data contents, much like including different forms of usage guidance with over-the-counter medication is needed to properly communicate how to use its contents. Independent variables included brief, cautionary verbiage in report footers, report-specific abstracts, and report-specific interpretation guides. The dependent variable was accuracy of data analysis-based responses. The researcher explored three data analysis supports provided by a data system, each framed in two different formats, by presenting 211 elementary and secondary educators in ethnically and culturally diverse southern California with different versions of the same two student achievement data report environments. Each of these report sets fit into one of the following treatment categories (a) no added analysis support; (b) analysis support by way of footers directly on the reports, which were offered in two different framing styles; (c) analysis support by way of abstracts, which accompanied the reports and were offered in two different framing styles; and (d) by way of interpretation guides, which accompanied the reports and were offered in two different framing styles (see *Appendix C* for reports and handouts). The researcher then compared the results of educators using data system reports embedded with data analysis guidance in the varied formats noted above (a-c). Participant responses were collected through a web-based questionnaire crafted and administered in Google Docs, taking advantage of the Google Form feature,

and involved groups of no more than 30 respondents at each administration time at each participant's school site. Data was collected at one point in time for each participant within a one-month research window. Findings from this research are suited to identify whether data systems used by educators can help prevent common analysis mistakes by providing analysis support within the interface and the reports they are used to generate.

This paper features an exploration of the concept of over-the-counter data: essentially, the prospect of improving educators' data use by embedding data usage guidance within the data systems they are using to analyze data, just as over-the-counter medication is packaged with usage guidelines. *Table 3.01* illustrates research questions that were used to explore the impact of three variables on data analysis accuracy: brief, cautionary verbiage in report footers; report abstracts; and interpretation guides. *Table 3.02* illustrates research questions relating to variables that could possibly have impacted educators' likelihood of using the investigated supports and/or educators' data analyses, and were thus included to help better understand the implications of findings addressed by the primary research questions illustrated in *Table 3.01*.

Research method and design will be discussed and will address the appropriateness of the method, use of a pilot test, and alignment with other study components. Participants and materials/instruments involved in the study will be explained. Variables and data procedures will be outlined, as will methodological assumptions, limitations, and delimitations. Finally, the paper will feature a thorough account of ethical assurances, followed by a summary.

Table 3.01: *Primary Research Questions and Hypotheses*

Research Question	Null Hypothesis	Alternative Hypothesis
Q1. What impact does data analysis guidance accompanying a data system report in the form of footer, abstract, or interpretation guide have on how frequently educators draw accurate conclusions concerning student achievement data?	**H1$_0$.** The null hypothesis was that accompanying a report with a support containing analysis guidance in the form of footer, abstract, or interpretation guide would not have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.	**H1$_a$.** The alternative hypothesis was that accompanying a report with a support containing analysis guidance in the form of footer, abstract, or interpretation guide would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.
Q2a. What impact does a footer with analysis guidelines on a data system report have on how frequently educators draw accurate conclusions	**H2a$_0$.** The null hypothesis was that accompanying a report with a supportive footer containing analysis guidance would not have a positive impact on the	**H2a$_a$.** The alternative hypothesis was that accompanying a report with a supportive footer would have a positive impact on the frequency of accurate

concerning student achievement data?	frequency of accurate conclusions educators drew concerning student achievement data.	conclusions educators drew concerning student achievement data.
Q2b. What impact does the manner in which a footer is framed, in terms of moderate differences in length and text color, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?	**$H2b_0$.** The null hypothesis was that the manner in which a footer was framed, in terms of moderate differences in length and text color, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.	**$H2b_a$.** The alternative hypothesis was that the manner in which a footer was framed, in terms of moderate differences in length and text color, would have an impact on the frequency of accurate conclusions educators drew concerning student achievement data.
Q3a. What impact does providing a report abstract, such as a one-page reference sheet with report purpose and data use warnings specific to the	**$H3a_0$.** The null hypothesis was that including a report abstract with a data system report would not have a positive impact on the frequency with which	**$H3a_a$.** The alternative hypothesis was that including a report abstract with a report would have a positive impact on the frequency of accurate

report it accompanies, with a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?	educators drew accurate conclusions concerning student achievement data.	conclusions educators drew concerning student achievement data.
Q3b. What impact does the manner in which an abstract is framed, in terms of moderate differences in density and header color, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?	**$H3b_0$.** The null hypothesis was that the manner in which an abstract was framed, in terms of moderate differences in density and header color, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.	**$H3b_a$.** The alternative hypothesis was that the manner in which an abstract was framed, in terms of moderate differences in density and header color, would have an impact on the frequency of accurate conclusions educators drew concerning student achievement data.
Q4a. What impact does providing an interpretation guide, such as a two-sided	**$H4a_0$.** The null hypothesis was that including an interpretation guide with a	**$H4a_a$.** The alternative hypothesis was that including an interpretation

reference sheet with analysis guidance and examples specific to the report it accompanies, with a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?	data system report would not have a positive impact on the frequency with which educators drew accurate conclusions concerning student achievement data.	guide with a report would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.
Q4b. What impact does the manner in which an interpretation guide is framed, in terms of moderate differences in length and information quantity, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?	**H4b$_0$.** The null hypothesis was that the manner in which an interpretation guide was framed, in terms of moderate differences in length and information quantity, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.	**H4b$_a$.** The alternative hypothesis was that the manner in which an interpretation guide was framed, in terms of moderate differences in length and information quantity, would have an impact on the frequency of accurate conclusions educators drew concerning student achievement data.

Table 3.02: *Secondary Research Questions Informing Implications Addressed by Primary Research Questions*

Research Question	Null Hypothesis	Alternative Hypothesis
Q5a. What impact does an educator's school site level type (i.e., elementary or secondary) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?	**$H5a_0$.** The null hypothesis was that an educator's school site level (i.e., elementary, middle/junior high, or high school) would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.	**$H5a_a$.** The alternative hypothesis was that an educator's school site level (i.e., elementary, middle/junior high, or high school) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
Q5b. What impact does an educator's school site level (i.e., elementary, middle/junior high, or high school) have on the frequency with which he or she draws accurate	**$H5b_0$.** The null hypothesis was that an educator's school site level type (i.e., elementary or secondary) would have an impact on the frequency of accurate conclusions he or she drew	**$H5b_a$.** The alternative hypothesis was that an educator's school site level type (i.e., elementary or secondary) would not have an impact on the frequency of accurate conclusions he

conclusions concerning student achievement data?	concerning student achievement data.	or she drew concerning student achievement data.
Q5c. What impact does an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, have on the frequency with which he or she draws accurate conclusions concerning student achievement data?	**H5c$_0$.** The null hypothesis was that an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.	**H5c$_a$.** The alternative hypothesis was that an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
Q5d. What impact does an educator's school site English Learner (EL) population have on the frequency with which he or	**H5d$_0$.** The null hypothesis was that an educator's school site English Learner (EL) population would have an impact on the	**H5d$_a$.** The alternative hypothesis was that an educator's school site English Learner (EL) population would not have

she draws accurate conclusions concerning student achievement data?	frequency of accurate conclusions he or she drew concerning student achievement data.	an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
Q5e. What impact does an educator's school site Socioeconomically Disadvantaged population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?	**H5e$_0$.** The null hypothesis was that an educator's school site Socioeconomically Disadvantaged population would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.	**H5e$_a$.** The alternative hypothesis was that an educator's school site Socioeconomically Disadvantaged population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
Q5f. What impact does an educators' school site Students with Disabilities population have on the frequency with which he or she draws accurate conclusions concerning	**H5f$_0$.** The null hypothesis was that an educator's school site Students with Disabilities population would have an impact on the frequency of accurate conclusions he or she drew	**H5f$_a$.** The alternative hypothesis was that an educator's school site Students with Disabilities population would not have an impact on the frequency of accurate conclusions he

student achievement data?	concerning student achievement data.	or she drew concerning student achievement data.
Q6a. What impact does an educator's veteran status have on the frequency with which he or she draws accurate conclusions concerning student achievement data?	**H6a$_0$.** The null hypothesis was that an educator's veteran status would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.	**H6a$_a$.** The alternative hypothesis was that an educator's veteran status would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
Q6b. What impact does an educator's current professional role (e.g., teacher, site/school administrator, etc.) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?	**H6b$_0$.** The null hypothesis was that an educator's current professional role (e.g., teacher, site/school administrator, etc.) would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.	**H6b$_a$.** The alternative hypothesis was that an educator's current professional role (e.g., teacher, site/school administrator, etc.) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Q6c. What impact does an educator's perception of his or her own data analysis proficiency impact the frequency with which he or she draws accuDatarate conclusions concerning student achievement data?	**$H6c_0$.** The null hypothesis was that an educator's perception of his or her own data analysis proficiency would be related to the frequency of accurate conclusions he or she drew concerning student achievement data.	**$H6c_a$.** The alternative hypothesis was that an educator's perception of his or her own data analysis proficiency would not be related to the frequency of accurate conclusions he or she drew concerning student achievement data.
Q6d. What impact does an educator's professional development over the past year, devoted specifically to *how* to analyze student data, have on the frequency with which he or she draws accurate conclusions concerning student achievement data?	**$H6d_0$.** The null hypothesis was that an educator's professional development over the past year, devoted specifically to how to analyze student data, would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.	**$H6d_a$.** The alternative hypothesis was that an educator's professional development over the past year, devoted specifically to how to analyze student data, would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Q6e. What impact does the number of graduate-level educational measurement courses an educator has taken have on the frequency with which he or she draws accurate conclusions concerning student achievement data?	**H6e$_0$.** The null hypothesis was that an educator's number of graduate-level educational measurement courses would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.	**H6e$_a$.** The alternative hypothesis was that an educator's number of graduate-level educational measurement courses would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Research Method and Design

An effective study of the potential of data analysis supports accompanying data system reports to increase users' analysis accuracy had to examine multiple reporting environments that could be replicated by a data system. First, the experimental quantitative study had to show educators make analysis errors when using typical data system reports, which do not contain analysis guidance (a) on the reports, or by way of supplemental documentation such as (b) abstracts or (c) interpretation guides that can be reached via link or provided with report printouts. The researcher then needed to compare those results to results for educators using data system reports embedded with data analysis guidance in the varied formats noted above (a-c). The research design also had to allow for framing influences by presenting each of the three data analysis supports (a-c) in two different formats. This allowed the study to measure not only whether – and to

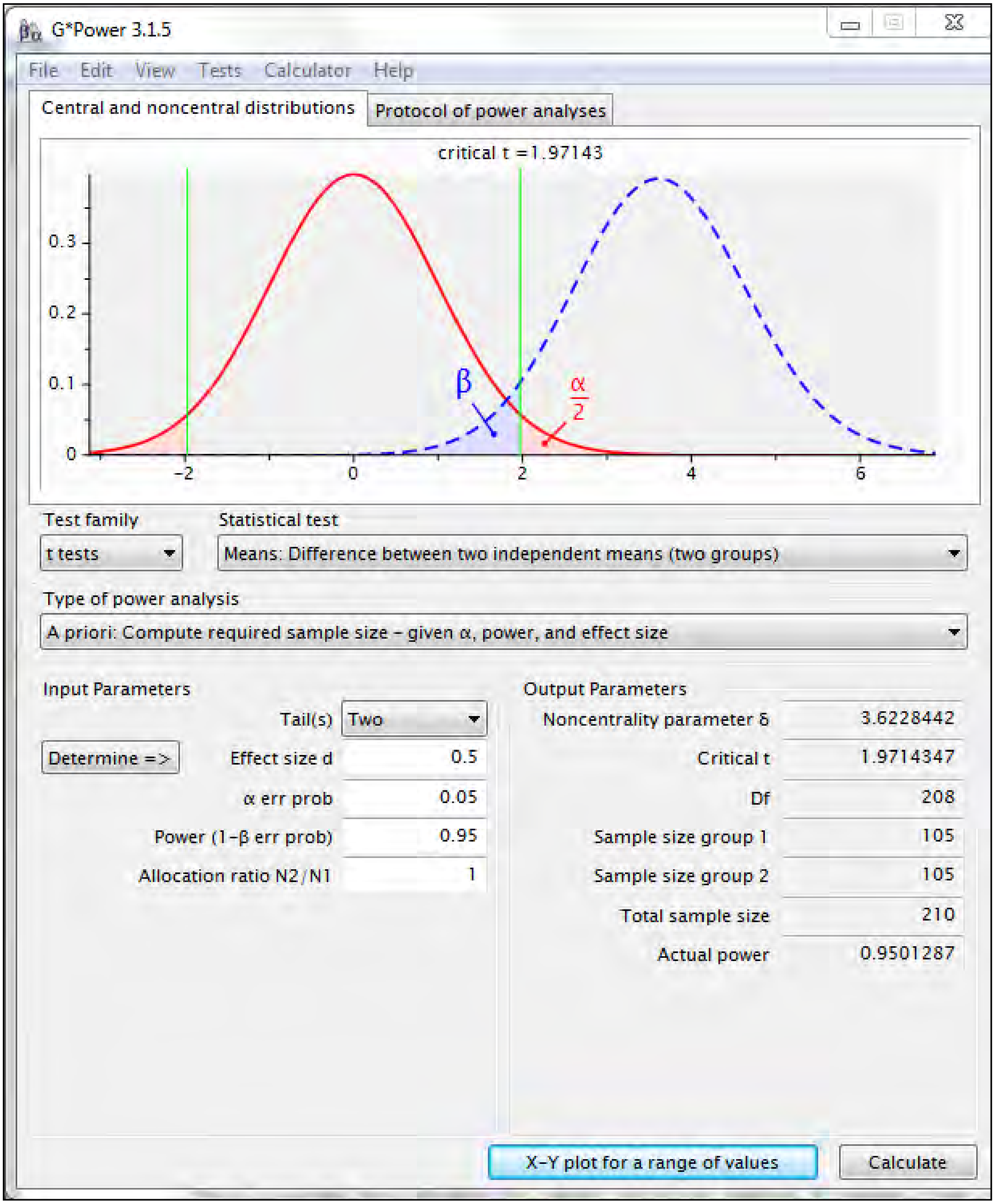


Figure 3.01: *Two-Tailed T-Test*

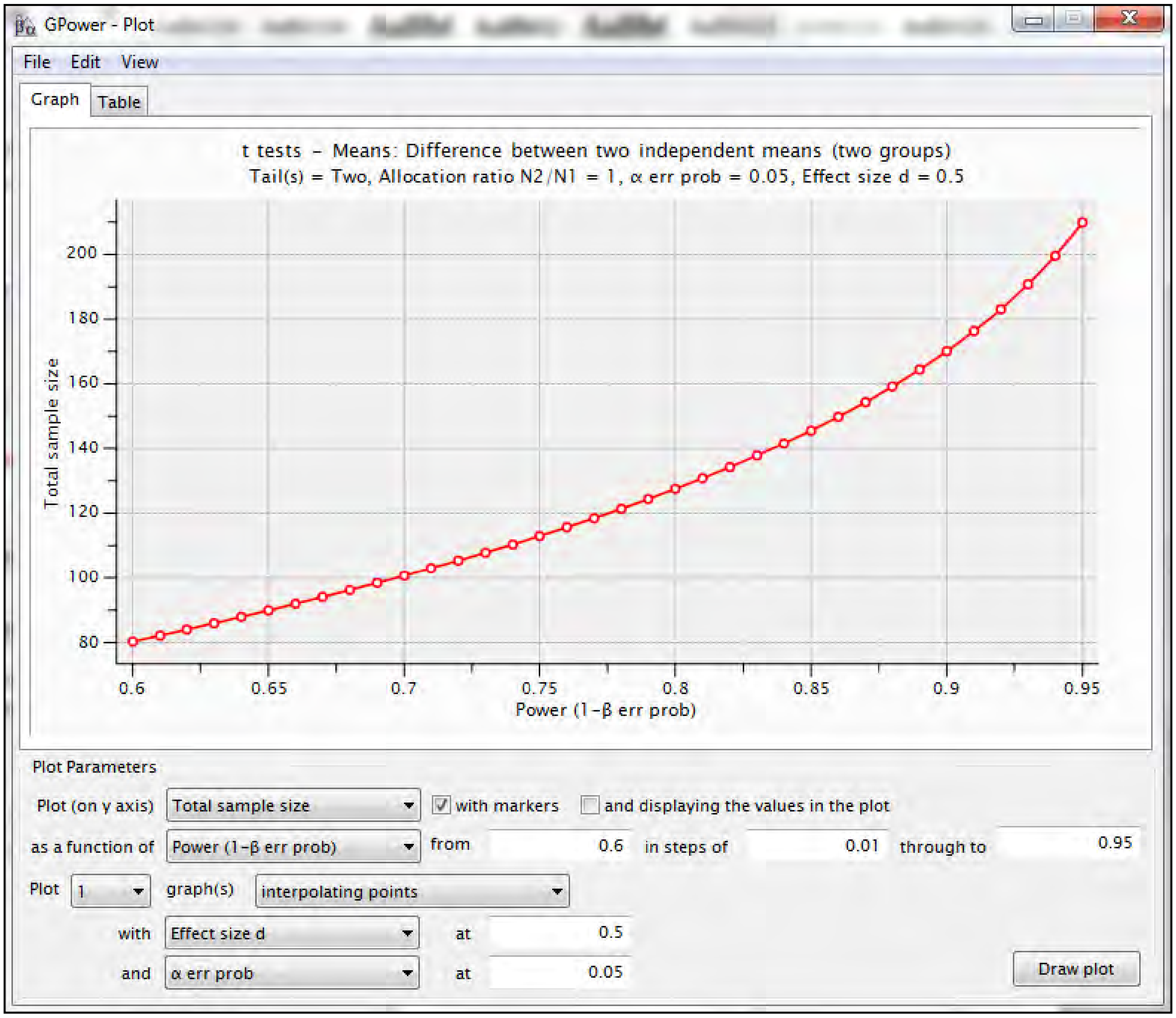


Figure 3.02: *Two-Tailed T-Test X-Y Plot Graph*

what extent – each analysis support can increase analysis accuracy, but also the more effective way in which to frame each support.

The G*Power 3.1 statistical analysis tool can be used to conduct a priori analysis, which involves calculating the necessary sample size by specifying values for the required significance level α, the desired statistical power 12β, and the population effect size that has yet to be determined (Faul, Erdfelder, Buchner, & Lang, 2009). To determine ideal sample size through priori power analysis, the researcher conducted a two-tailed t-test calculating the difference between two independent means utilizing the

G*Power 3.1 statistical analysis tool. For this analysis's details, see *Figure 3.01* for all input and output parameters and see *Figure 3.02* for an X-Y plot graph showing the power (1- β probability of correctly rejecting the null hypothesis) in relation to sample size. Input parameters included: tails = two, effect size d = 0.5, α error of probability (alpha, the probability of a type I error) = 0.05, and power (1-β error of probability for a type II error) = 0.95. Output parameters included noncentrality parameter δ = 3.6228442, critical t = 1.9714347, Df = 208, sample size group1 = 105, sample size group 2 = 105, total sample size, = 210, actual power = 0.9501287. The priori two-tailed t-test thus resulted in a recommended sample size of at least 210 educators.

However, the researcher also conducted an F-test linear multiple regression analysis, fixed model, R^2 deviation from zero, using the G*Power 3.1 statistical analysis tool. For this analysis's details, see *Figure 3.03* for all input and output parameters and see *Figure 3.04* for an X-Y plot graph showing the power (1- β probability of correctly rejecting the null hypothesis) in relation to sample size. Input parameters included: effect size f^2 = 0.15, α error of probability (alpha, the probability of a type I error) = 0.05, power (1-β error of probability for a type II error) = 0.95, and number of predictors based on independent variables =7. Output parameters included noncentrality parameter λ = 22.9500000, critical F = 2.0732820, numerator df = 7, denominator df = 145, total sample size = 153, and actual power = 0.9503254. The priori F-test thus resulted in a recommended sample size of at least 153 educators. However, since the 210 sample size resulting from the two-tailed t-test was greater than 153, responses from 211 participants were collected for the study in order to exceed even the more rigorous recommendation. See the *Chapter 3: Research Method: Research Method and Design: Regression analysis*

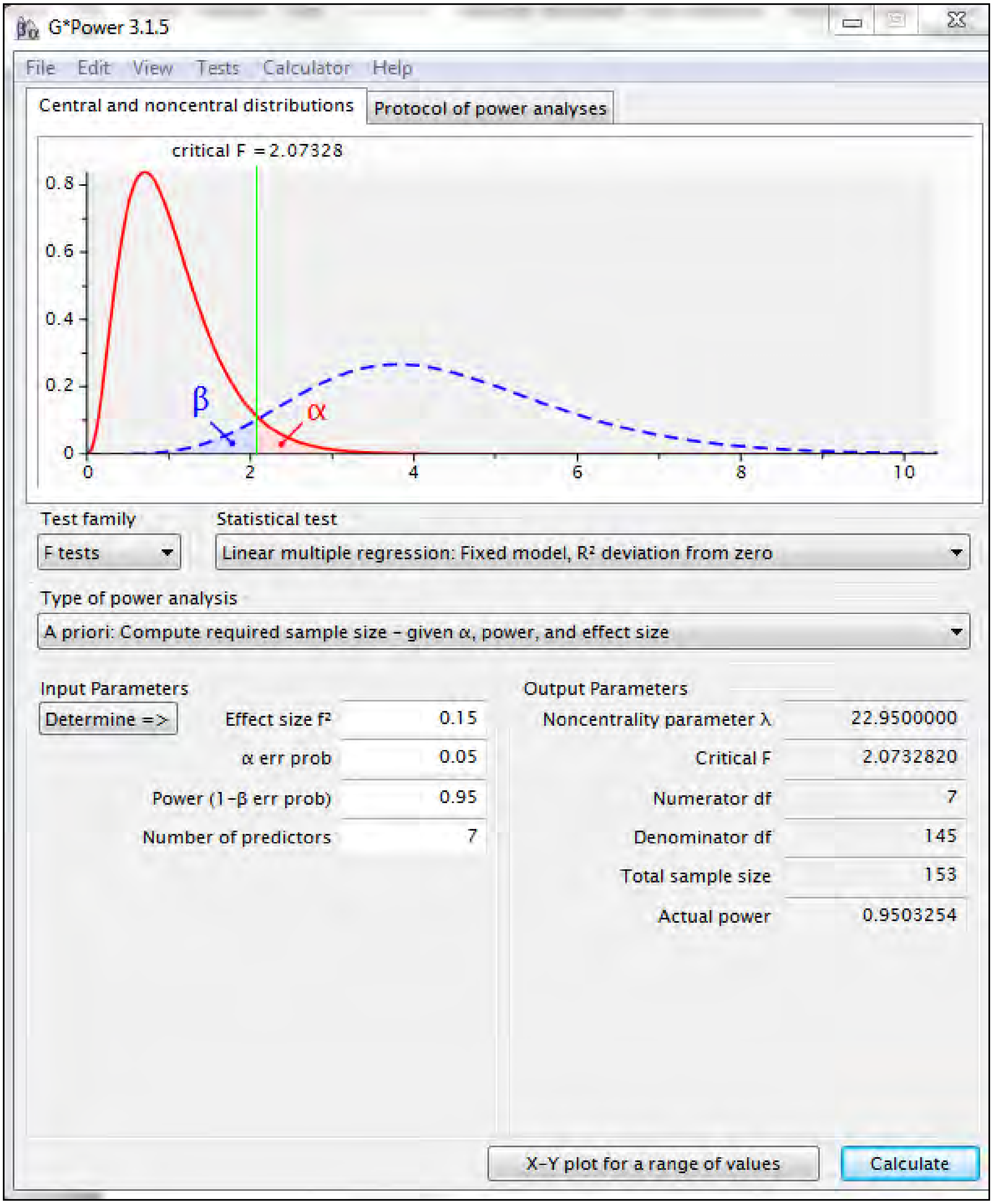


Figure 3.03: *F-Test*

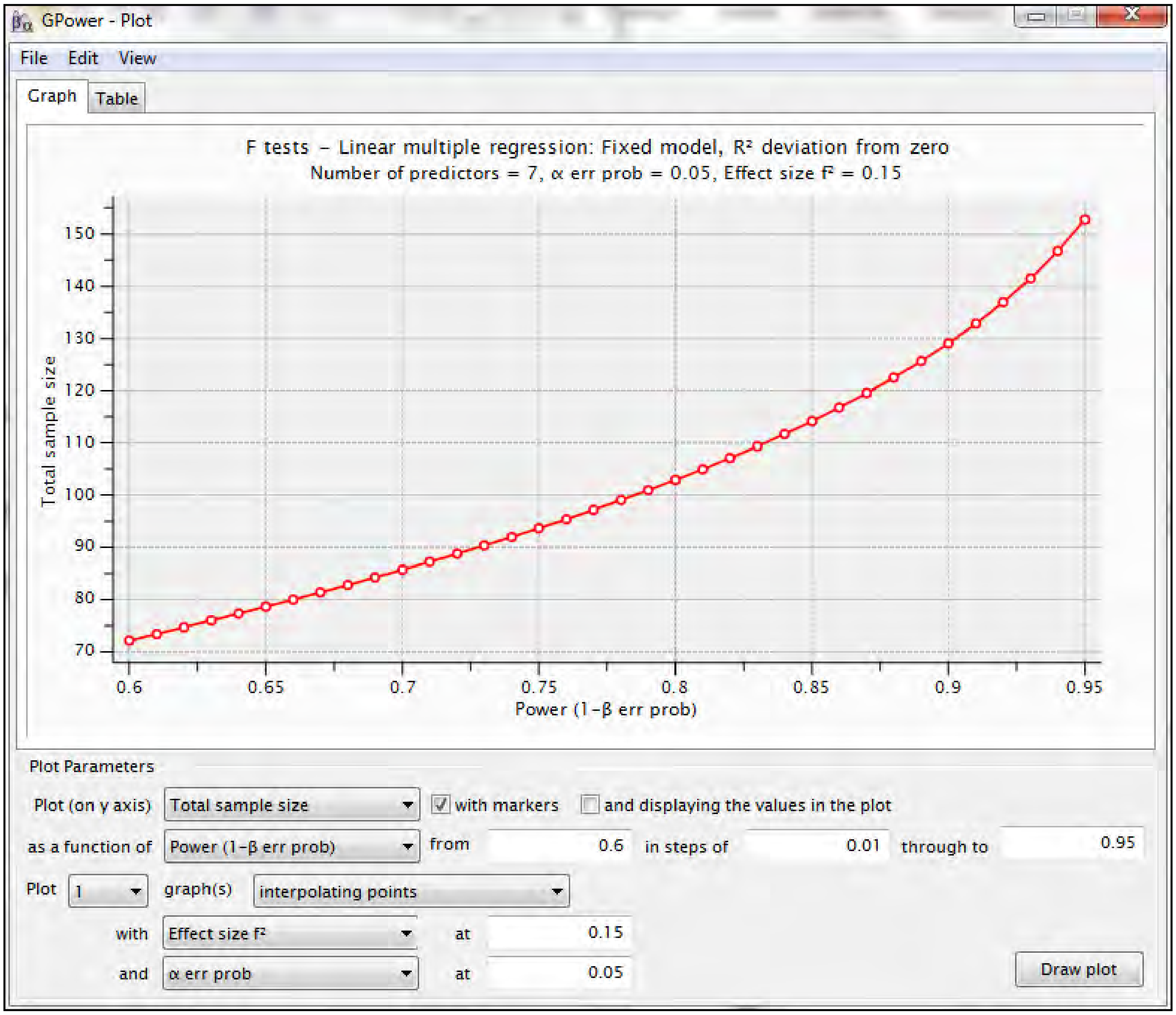


Figure 3.04: *F-Test X-Y Plot Graph*

section for details on the regression analyses that was also applied.

To avoid threats to external validity, the researcher needed to avoid interaction of selection and treatment in this approach. Thus the study included educators from varied school sites and of varied roles, such as elementary level and secondary level, veterans and non-veterans, teachers and administrators, etc. See *Table 3.03* and *Table 3.04* for the number and percent of participants within each characteristic category. The study also employed a random, cross-sectional sampling procedure. The 211-participant size

provided a more reliable data sampling, the mix assisted in the stratification of the population, and the randomization offered the ability to generalize to the education population at large.

Table 3.03: *Participant Site Characteristics*

	Participants	
Category	*n*	%
County		
Los Angeles Unified School District	32	15%
Orange	42	20%
San Bernardino	137	65%
School District		
Alta Loma School District	16	8%
Buena Park School District	42	20%
Chino Valley Unified School District	11	5%
Etiwanda School District	31	15%
Mountain View School District	79	37%
Los Angeles Unified School District	32	15%

School Site		
Buena Park Junior High	14	7%
Charles G. Emery Elementary	28	13%
Creek View Elementary	22	10%
Etiwanda Colony Elementary	31	15%
Grace Yokely Middle	33	16%
Hermosa Elementary	16	8%
Ranch View Elementary	24	11%
Rolling Ridge Elementary	11	5%
Sylmar High	32	15%
School Level Type		
Elementary	132	63%
Secondary	79	37%

School Level		
Elementary	132	63%
Middle/Junior High	47	22%
High School	32	15%
2012 Growth Academic Performance Index (API), California State Accountability Measure Ranging 200-1000		
677	32	15%
794	33	16%
815	24	11%
827	14	7%
847	22	10%
891	28	13%
893	16	8%
895	31	15%
916	11	5%

% of Site's Students Who Are English Learners (29% Mean)		
8%	16	8%
10%	31	15%
16%	11	5%
27%	22	10%
30%	33	16%
33%	24	11%
38%	32	15%
45%	14	7%
46%	28	13%

% of Site's Students Who Are Socioeconomically Disadvantaged (52% Mean)		
22%	11	5%
23%	31	15%
31%	16	8%
43%	28	13%
56%	22	10%
61%	57	27%
78%	46	22%
% of Site's Students with Disabilities (10% Mean)		
5%	16	8%
8%	28	13%
9%	38	18%
10%	33	16%
11%	33	16%
12%	32	15%
13%	31	15%

Table 3.04: *Participant Characteristics*

	Participants	
Category	*n*	%
Veteran Status: Length of Time Working as an Educator (e.g., Teacher or Administrator) for Students under 19 Years of Age		
Less than 1 Year	2	1%
Minimum of 5 Years	20	9%
Minimum of 10 Years	33	16%
Minimum of 15 Years	67	32%
Minimum of 20 Years	89	42%
Role: Best Description of Current Position		
Teacher	199	94%
Colleague Coach (e.g., Teacher on Special Assignment)	2	1%
Site/School Administrator	8	4%
District Administrator	2	1%

Perceived Proficiency at Analyzing Student Performance Data		
Very Proficient	45	21%
Somewhat Proficient	139	66%
Not Proficient	22	10%
Far from Proficient	5	2%
Professional Development Obtained within Past Year, Specifically Focused on Learning How to Correctly Interpret Student Data		
0 Hours	87	41%
Minimum of 1 Hour	48	23%
Minimum of 2 Hours	39	18%
Minimum of 5 Hours	19	9%
Minimum of 8 Hours	18	9%

Graduate-Level Courses Taken, Specifically Dedicated to Educational Measurement		
0 Courses	100	47%
Minimum of 1 Course	51	24%
Minimum of 2 Courses	35	17%
Minimum of 3 Courses	11	5%
Minimum of 4 Courses	14	7%

The researcher collected response data through a web-based, self-administered Google Docs survey form that allows for efficient collection without initial interpretation. However, the researcher was present with participants during the survey completion process in case clarification was needed on how to proceed. If no one were physically present to oversee the survey's completion, the study would not have allowed participants to ask questions and receive clarification on the survey process, which would mean potential weakness for the study.

Appropriateness of method. The quantitative survey method lent itself well to this study, as it explored performance on data questions with clear answers, as are regularly encountered by teachers seeking to understand student data. The non-subjective nature of these questions matched a quantitative study, and the survey format allowed response data to be collected efficiently and in a way that required no initial interpretation – and thus minimal risk of misunderstanding or accidental alteration – by the researcher.

The economy of the design allowed the study to incorporate a larger – and thus more reliable – sample size of data.

Behavioral Economics. This study related to improving the accuracy of educators' data analyses, as enacted in the thought portion – or "data-informed" portion – of data-informed decision-making. The process of thinking and deciding is influenced by behavioral economics facets such as priming, biases, heuristics, prototypes, judgments, anchoring, and framing (Kahneman, 2011). Thus data-informed thoughts are believed to influence decision-making. For example, even small and seemingly insignificant differences in how content is arranged can mean a significant difference in the decisions people make based on that content (Thaler & Sunstein, 2008). This study covered reports and supplemental documentation that can be generated from within the environment of the online data system, as the study's purpose lay in finding ways data systems can be improved to facilitate improved data analyses. Thus conditions outside of those that can be controlled within a data system were not manipulated or used as variables in the study. Nonetheless, behavioral economics research still influenced this study's design (see *Chapter 2: Literature Review: Behavioral Economics and Data-Informed Decision-Making* for descriptions of the behavioral economics dimensions and ways in which behavioral economics also influences data-informed decision-making).

Priming. Priming is a dimension of behavioral economics that involves one idea resulting in another, among many. Basically, a subtle influence such as a hint of an idea primes one's thoughts, which then impact one's actions in ways that can be surprisingly significant (Thaler & Sunstein, 2008). When applied to data-informed decision-making, an important source of priming can involve resources educators interact with before

viewing data to inform their decisions: the resources prime the educator's thoughts concerning the data, and then those thoughts prime the educator's decisions. For example, Goodman, and Hambleton (2004) noted the value gained in states that accompanied data reports with information for parents to read *before* reading and interpreting the data. More subconscious priming sources involved in data-informed decision-making include the environment in which analyses take place and the individual's associations with those facilitating the session, the data system used and the individual's associations with technology, etc.

While the behavioral economics concept of priming were applied to this study's design as some participants were presented with resources they could choose to review before analyzing data system reports, awareness of more subconscious priming sources also impacted study precautions. For example, due to biases and feelings participants may attach to members of staff, the researcher was the clear facilitator of the study session. Since the researcher was not a colleague of the respondents, this helped to prevent their negative, positive, or otherwise biased feelings concerning coworkers from influencing their analyses.

Likewise, many educators are intimidated by technology (Combs, 2004; Rodriguez, 2008) or do not use technology as much as others. For example, only 44% of educators who have access to data systems use them directly rather than only reading printed versions of reports *others* use the data systems to generate (Underwood, Zapata-Rivera, & VanWinkle, 2008). This is one reason why study participants interacted with printed versions of reports rather than online versions that require use of technology during analysis. Since some educators are also intimidated by data (Underwood et al.,

2010), reports used in the study also conformed to research-based recommendations concerning the exclusion of intimidating features like jargon and statistical terms that could have a negative priming effect (see *Chapter 2: Literature Review: History of Specific Research Contributions* for a historic timeline of research-based recommendations for improving report design).

Kahneman (2011) also found although priming people with thoughts of money results in more independence and determination to solve problems, it also results in selfishness and resistance to help others. Since data-driven decision-making in education is founded on the goal of helping students succeed, involving money – such as payment for time spent analyzing data or participating in the study – could be detrimental to participants' performance on the data analysis questions related to data-driven decision-making for students. This study did not involve any monetary compensation for participation.

Biases, heuristics, prototypes, and judgments. Research confirms that decisions people make are inherently flawed due to factors such as bias (Thaler & Sunstein, 2008). For example, the institutional nature of military decision-making processes (MDMP), organizational culture, and individuality all impact the heuristics and biases that influence how military commanders respond to surprises while in action (Williams, 2010). System 1 thought processes use biases and heuristics, such as prototypes, to speed up thinking and decision-making; a social example of this is a stereotype, which does not necessarily lead to an accurate conclusion (Kahneman, 2011). Biases, heuristics, and prototypes, as well as the judgments to which they lead, are not always undesirable, but they can cause flawed judgments, such as where data-informed decision-making is concerned. For

example, teachers' biases impact conclusions they draw when making data-informed decisions (Park, 2008).

Just as biases, heuristics, and prototypes impact data-informed decision-making, consideration of them impacted the design of this data-related study. For example, educators generally possess some level of racial bias (Day, 2010). Thus this study's data analysis questions and the reports on which the questions were based involved no subgroup comparisons, such as comparisons between ethnicities, races, or other demographics-based groups. Likewise, the questions and reports related to student-level data identified students as Student A, Student B, etc. as opposed to using actual names, which could be associated with particular ethnicities, races, or socio-economic status. Whenever a set or category is homogeneous enough to have a prototype, the brain will automatically access the prototype to consider the mean values associated with its members when making a decision (Kahneman, 2003). Avoiding questions and reports related to categories commonly associated with prototypes will help to prevent biases and heuristics from skewing respondents' analyses.

Biases and judgments also constitute a key reason behind the necessity of this study. One might think accompanying a data report with a footer, abstract, or interpretation guide would invariably increase the accuracy of the user's analyses. However, behavioral economics tells us this is not a guaranteed phenomenon. If the concept of rational analysis that was popular before the rise of behavioral economics held sway, educated people presented with a clear explanation for how to find a correct answer would always find the correct answer. However, research on data reporting supports the behavioral economics premise that people do not always behave most

effectively. For example, too much information or text on reports can overwhelm users and actually cause them to make mistakes (Hattie, 2010; VanWinkle et al., 2011; Zapata-Rivera & VanWinkle, 2010). In addition, social science research from the last 40 years confirms that people do not always make the best decisions or do what is perceived as best, even when those decisions or actions directly impact their wellbeing (Thaler & Sunstein, 2008). Thus questions remain concerning the best ways to assist data analysis accuracy within data systems and their reports.

Anchoring. The anchor heuristic is a value someone considers before estimating the quantity of something, and anchoring effect is the phenomenon that causes his or her estimate to stay closer to the anchor than it might have been is the anchor were not considered (Kahneman, 2011; Thaler & Sunstein, 2008). Anchoring usually results in an inaccurate estimate (Williams, 2010). Anchoring can occur in data-informed decision-making when educators have preconceived notions of an entity's performance. Essentially, the anchor can prime the teacher's thoughts, which then prime his or her actions.

Research on anchoring influenced the manner in which this study's survey was designed. All data analysis questions on the survey were devoid of statistical numbers. Recommended uses of data in educational settings call for a focus on the relationship between practice and desired outcomes (Bernhardt, 2007). Analyses and the information they produce increase in quality when multiple measures are compared (Bernhardt, 2004). For example, "Are at least 57% of the students scoring Proficient" is a question likely to be hampered by anchoring effect. However, it includes no comparison of multiple measures, and regardless of the answer it gives no concrete direction concerning

educational practice. Conversely, data analysis questions on this study's survey involved the comparison of multiple measures and rendered specific information concerning practice in the hypothetical educational settings. Thus the questions not only avoided anchoring heuristics, but they also mirrored the types of questions educators should be asking when analyzing their own data from non-hypothetical reports conveying similar data.

Framing. Framing applies to the presentation of information, and presenting the same information to someone in different ways will often result in different emotions and different levels of difficulty in understanding or analyzing the information (Kahneman, 2003, 2011). The manner in which content is organized for people using it to make decisions significantly impacts those decisions (Thaler & Sunstein, 2008). Framing thus plays a large role in data analysis accuracy and data-informed decision-making (see *Chapter 2: Literature Review: History of Specific Research Contributions* for a historic timeline of research-based recommendations for report design, which relate to framing). The reports used in this study subscribed to leading research-based recommendations concerning the best ways in which to frame the data in report format, though they did so in a way that did not deviate from what is commonly seen in data systems currently on the market. In other words, reports used in the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study adhered to the better data presentations commonly seen in data systems, but they did not adhere to the best data presentations that – despite being more effective – are not yet commonly seen in student data systems.

Suggested ways to present analysis guidance in footers, abstracts, and interpretation guides were utilized in this study, but the best manner in which to frame

these resources had not yet been determined in regards to direct impact on analysis accuracy. Thus each of the three support resources used in this study were framed in two different formats for respondents.

Research on framing also influenced the manner in which this study's survey was designed. On multiple choice questions, an option receives a large advantage if it is presented as the default (Johnson, Hershey, Meszaros, & Kunreuther, 1993). Thus the data analysis survey questions used in this study presented no distractors or answers as defaults; rather, no answer options were preselected, meaning respondents had to select one of the equally-presented answers before continuing to the next question. All survey questions were multiple choice, nominal, close-ended questions.

Framing was also a key reason behind the necessity of this study. While research already existed concerning the best data displays to use to improve educators' analyses, the best way in which to frame analysis support within a data system to improve educators' analyses had not yet been determined. This gap in research literature was one the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study was designed to fill.

Regression analysis. Linear regression analysis and multiple linear regression analysis were both used to investigate the relationship between the study's dependent and independent variables. See *Chapter 3: Research Method: Data Collection, Processing, and Analysis: Regression Analysis* for regression analysis details. This includes a table illustrating the survey's research question variables with the corresponding regression analysis features designed to address them.

Pilot test. The researcher conducted an initial pilot test with five educators to garner feedback, both from their answers using the instrument and from their verbal feedback on the instrument itself and the time it took to complete the survey. This was done in order to improve the questions and format prior to the survey's official administration, though no adjustments were necessary. These educators were representative of the varied roles and backgrounds of the educators who ultimately served as participants in the study, and the materials they were given varied in order to test all variable types. See *Table 3.05* for pilot test participant details, reporting environments, and survey completion time. No participants took longer than 15 minutes to complete the survey, and feedback suggested no part of the survey was confusing or warranted changing.

Table 3.05: *Pilot Test Participants, Materials, and Survey Completion Time*

Participant Experience	Report Materials	Time
Participant A had been a middle school teacher, middle school technology coordinator, middle school site administrator (assistant principal), and county department of education administrator	• Report with No Footer (Plain Report) • No Abstract • No Interpretation Guide	10 min.

Participant B was an elementary school teacher and had been an elementary school site administrator (assistant principal and principal)	• Report with Footer A (Longer) • No Abstract • No Interpretation Guide	15 min.
Participant C was a high school teacher and had been a junior high school teacher	• Report with Footer B (Shorter) • No Abstract • No Interpretation Guide	7 min.
Participant D was a junior high school teacher	• Report with No Footer (Plain Report) • Abstract B (Less Dense) • No Interpretation Guide	10 min.
Participant E was a junior high school site administrator (assistant principal) and had been an elementary school teacher	• Report with No Footer (Plain Report) • No Abstract • Interpretation Guide A (3 Pages)	15 min.

Alignment with other study components. The survey questions were constructed with the aim of assessing the accuracy with which educators draw inferences when viewing student performance data contained in report formats typical of most data systems versus reports containing or accompanied by some level of data analysis support. The researcher administered the survey in 10 sessions in computer labs at nine school

sites, as one school site was visited twice to administer the survey to two separate groups on two separate days. The researcher passed out a copy of the Informed Consent Form, which was approved by Northcentral University's Institutional Review Board (IRB), to each participant as he or she arrived at the computer lab so all attendees would have time to read it. The researcher then briefly introduced the study so participants knew key facts such as the nature of the study, the anonymity of responses, participation was voluntary, there were no benefits to participating other than contributing to field literature in a way that was hoped to eventually help educators and students, and there were no penalties of any kind if any attendees wished not to participate. All attendees in all cases opted to sign the Informed Consent Form and participate.

After the researcher collected all Informed Consent Forms, the researcher handed each participant a different folder containing reports and handouts to read in conjunction with survey questions, but not all participants will receive the same reports or handouts. See *Table 3.06* for an indication of what each folder contained. Seven different folder colors were used, and these were stacked ahead of time in the alternating format of white, yellow, green, blue, purple, red, black, then white again, etc. so they were distributed evenly with participants seated as far as possible from participants with the same folder contents.

The researcher called participants' attention to the stickers on the folder covers that stated their color to accommodate color blind participants, as folder color was used to determine which version of Question 8 each participant answered on the survey. The researcher also pointed out the two stickers on folders' two inside pockets that indicated which materials related to Report 1 and should be used to answer survey Questions 4 and

5, and which materials related to Report 2 and should be used to answer survey Questions 6 and 7. The researcher then prompted participants to begin the online survey on the computer, which was set to a web address that could not be accessed by anyone who did not have the exact uniform resource locator (URL), which was changed after each survey administration. The survey prompted participants to hold their folders in the air as they finished. This allowed the researcher to check each computer screen to ensure the survey was successfully submitted, which was only possible if all questions were answered since the required question setting was used on every survey question, and participants were exited from the online survey environment.

After results from all 211 participants were collected, the researcher used straightforward categorical scales in the form of correct/incorrect for data analysis questions, as the answers were clearly right or wrong based on the guidelines from the performance data's governing body (California Department of Education). For example, teachers viewing California Standards Test (CST) content cluster data from the state Standardized Testing and Reporting (STAR) Program were asked, "Which content cluster is most likely the school site's weakness?" The answer was clearly one of the five clusters, as indicated by the California Department of Education's *California Standardized Testing and Reporting Post-Test Guide Technical Information for STAR District and Test Site Coordinators and Research Specialists* (CDE, 2012b) and *California Standards Tests Technical Report* (CDE, 2011). The phrasing "most likely" also avoids the impact questions of significance would otherwise have on the question's answer. Because this single-correct-answer-per-question approach to data collection was objective and the study was quantitative, the need to scale respondents' analysis

responses further was circumvented and the percent of analysis-related questions participants answered correctly was used, both overall and in relation to whether or not a data analysis support related to the question was used by the respondent. The same would not be true if this had been a qualitative study.

Population

The study's population is comprised of public educators of all TK-12 school levels in the United States of America. There are approximately 3,250,600 public school teachers educating 47,315,700 students at 88,113 public schools, along with 57,000 instructional coordinators and supervisors (Strizek, Pittsonberger, Riordan, Lyter, & Orlofsky, 2006), totaling 3,307,600 U.S. educators. Such educators are of varied veteran levels, working in varied roles, and at schools with a range of demographics, such as high versus low performing and varied student populations. For example, at school sites where public educators are based nationwide, 10% of students are considered English Learners (EL), 23.4%-32% of students are considered Socioeconomically Disadvantaged depending on which indicator was used, and 13% are Students with Disabilities (U.S. Department of Education Institute of Education Sciences National Center for Education Statistics [USDEIESNCES], 2012).

Other population characteristics include:

- highly skilled: e.g., 95% of teachers are considered "highly qualified" by No Child Left Behind (NCLB) standards (American Institutes for Research [AIR], 2013), though there is debate concerning this label's merit.

- well-educated: e.g., 99% of American teachers have bachelor's degrees, 48% have master's degrees, and over 7% have more advanced graduate degrees (Papay, Harvard Graduate School of Education, 2007).
- embracing data use: e.g., most educators are eager to analyze and then act on the data they see (Hattie, 2010; van der Meij, 2008).

Sample

The procedure for sampling study participants was random and cross-sectional, incorporating responses from 211 educators of all TK-12 school levels to allow for the inclusion of all veteran levels, working in varied roles, and at schools with a range of demographics, such as high versus low performing and varied student populations. These demographic variables – accounted for in *Table 3.07*, *Table 3.08*, and *Table 3.09* – were included in the study's multiple regression analysis. The mix assisted in the stratification of the population, and the 211-sample size was possible using a computer survey collection of responses in order to garner a more reliable data sampling than would be possible with smaller numbers. Also, the larger sample number allowed for a better cross-sectional sampling.

This approach to involving varied participants from varied school sites allowed the sample drawn from the population to appropriately represent the actual educator population. For example, the sample involved participants representing all veteran levels, all credentialed educator roles, all perceived data analysis proficiency levels, all data analysis professional development categories, and all graduate-level educational measurement course categories. Likewise, the sites at which the sample participants were based represented the varied demographics of those nationwide. For example, 10% of

students in the United States are considered English Learners (EL), 23.4%-32% of students are considered Socioeconomically Disadvantaged depending on which indicator was used, and 13% are Students with Disabilities (USDEIESNCES, 2012). The sites at which this study's participants were based represented a spectrum encompassing these national statistics in all cases: sites ranging from 8% to 46% EL with a per-participant mean of 29% encompassed the national statistic of 10% EL, 22% to 78% Socioeconomically Disadvantaged with a per-participant mean of 52% encompassed the national statistic of 23.4%-32% Socioeconomically Disadvantaged, and 5% to 13% Students with Disabilities with a per-participant mean of 10% encompassed the national statistic of 13% Students with Disabilities. In terms of the academic achievement of students at school sites, the state of California's state accountability measure, which ranges from 200-1000 and is also used as a factor in federal accountability, is the Growth Academic Performance Index (API). The state average for 2012 was 788 Growth API (California Department of Education Analysis, Measurement, & Accountability Reporting Division, 2013). The sites at which this study's participants were based represented a spectrum encompassing this national statistics: sites ranging from 677 to 916 Growth API with a per-participant mean of 828 encompassed the national statistic of 788 Growth API. See *Table 3.03* and *Table 3.04* for the distribution of participant and site variables.

Initial subject recruiting did not begin until after approval was obtained from Northcentral University's Institutional Review Board (IRB) Committee. After that point in time, the researcher extended an invitation to participate in the study, as well as proposal guidelines, to 91 educators in southern California public school districts of

varied socioeconomic student populations. The researcher also created a website (www.overthecounterdata.com) that housed the study invitation for anyone to see at www.overthecounterdata.com/study. The independent information resource and community known as EdSurge (www.edsurge.com), which reaches educators, vendors, and others involved in educational technology, was kind enough to include mention of the study opportunity in EdSurge Newsletter #114 (April 17, 2013), Newsletter #115 (April 24, 2013), and Newsletter #116 (May 1, 2013). Thus the researcher took steps to extend the invitation to as many educators as possible. Conversations with interested parties followed to discuss specifics and ethical assurances.

Data was collected at one point in time for each participant within a 32-day research window of April 8, 2013, to May 10, 2013. Though there were efforts to select from a range of school demographics, the procedure for sampling these individuals at those sites was random. This randomization offered the ability to generalize results to educator populations. Each group survey session was set up by a school administrator at the site. Participation at each site was voluntary.

Materials/Instruments

Participant responses were collected through a web-based survey crafted and administered in Google Docs, employing the Google Form feature. Since an appropriate survey did not already exist, one was created specifically for this study (see *Appendix B*). The survey included 10 numbered questions and an additional, unnumbered question that impacted which version of Question 8 each respondent was asked. The Google Docs Form tool automatically assigned an anonymous ID to each respondent's data, which was

used in complete absence of participant names or employee numbers. The data was automatically, securely stored and password-protected online as soon as it was entered.

Response data was exported into Microsoft Excel® in order to be coded there and used with the Microsoft 2010 Data Analysis feature, and also used with Predictive Analytics Software (PASW) Version 18 with the Statistical Package for the Social Sciences (SPSS) Data Access Pack. After the response data was exported and saved on the researcher's password-protected computer, it was deleted from its online, Google Docs, password-protected environment in order to maximize security. Results were analyzed to (a) answer the study's seven primary research questions with related hypothesis strands, (b) answer the study's 11 secondary research questions with related hypothesis strands that served the sole role of informing implications addressed by the primary research questions, and (c) identify themes, patterns, relationships, and implications. This involved establishing categories and subcategories based on results and using the codebook mentioned below. In order to identify problems with typical data system report environments, results from reports that offered no analysis assistance were compared to results for educators using reports and resources that can come from data systems embedded with data analysis guidance in varied formats. Results were tabled, graphed to check for normal distribution, and tested to see if they were considered statistically significant. A descriptive analysis containing the means, standard deviations, and score ranges was then prepared in relation to the independent and dependent variables. See *Chapter 4: Results* for details such as the varying significance levels (p) used for different types of research questions.

The Google Docs Form "required question" setting was assigned to each survey question to eliminate the risk of response bias resulting from nonresponses on the survey. Though they could voluntarily stop at any time and were informed of this right, respondents were electronically forced to answer all questions in a group before proceeding to the next set, and thus it was impossible to complete the survey without answering all survey questions. The researcher checked each participant's computer screen after survey completion to ensure the survey was successfully finished and submitted, which it was in all cases. No participants failed to complete the entire survey. Descriptive information describing respondents and non-respondents would have been used in the event that some participants did not complete their surveys, but this was not necessary since all participants finished the survey.

Instrument. Please see *Appendix B* for printed copies of the pages from the actual, online Google Docs Form survey that was used for the study. All participants completed this survey. However, note that Question 8 was automatically individualized as one of four versions based on how respondents' folder colors tied to their versions of report and handout contents, are entered. In other words, the survey featured four different versions of Question 8, which was specific to the type of analysis support each respondent received. Thus Question 8 is featured on four pages of *Appendix B*, whereas each respondent only saw and responded to one of these four pages.

All analysis survey questions concerned data from state assessments with which the Californian study participants were most likely to be familiar with analyzing. One of these assessments was the California Standards Test (CST), as this assessment of student performance constituted the largest component of California's Standardized Testing and

Reporting (STAR) Program, which began in 1998, at the time this study was conducted. The CST was considered the educator participants' highest stakes test for both state and federal accountability. The other assessment was the California English Language Development Test (CELDT), which began in 2001. At the time this study was conducted, all Californian educators were supposed to consider a student's CELDT results when determining whether or not to recommend the English Learner (EL) for reclassification; no other assessment in these educators' state of California could be substituted for EL reclassification consideration. Thus:

- all study participants were expected, within the requirements of their professions, to be familiar with the assessments that generated the data participants analyzed in the study,
- the survey's data analysis questions were common questions all study participants were expected, within the requirements of their professions, to be familiar with, as they must answer such questions on an ongoing basis in relation to the same assessments that were used,
- the reports to which respondents referred to answer survey questions were typical of those Californian educators acquire from data systems, as they catered to the common assessments and questions noted above (see *Operational Definitions of Variables: Behavioral economics' impact on variables: Framing* for other ways in which the reports were typical of data system reports).

Instrument Validity and Reliability Concerns. The web-based survey through which participant responses were collected was crafted specifically for this study. This

was necessary because an instrument measuring data analysis accuracy of results from assessments with which all participants were most likely to be familiar – the CST and the CELDT, explained above – did not previously exist. Fortunately, validity and reliability concerns were circumvented as follows:

- each analysis question had one correct answer, and thus question distractors included all possible answers rather than having to be selectively determined;
- each analysis question's answer was objective rather than subjective, and thus there was no need for interpretation on the appropriateness of answers; and
- each analysis question and answer were based on straightforward guidelines published by the California Department of Education (CDE) to accompany each of the two state assessments and guide educators in the correct ways to draw conclusions from the data.

The CDE guidelines to which the survey's CST analysis questions conformed were featured in *California Standardized Testing and Reporting Post-Test Guide Technical Information for STAR District and Test Site Coordinators and Research Specialists* (CDE, 2012) and *California Standards Tests Technical Report* (CDE, 2011). The CDE guidelines to which the survey's CELDT analysis questions conformed were featured in *2011–12 Accountability Progress Reporting System: 2011–12 Title III Accountability Report Information Guide* (CDE, 2012a). These two resources were the most recent editions available at the time of this study. The *Chapter 3: Research Method: Materials/Instruments: Triangulation* section of this paper further explains considerations that were incorporated into the study handouts and survey questions to better facilitate triangulation.

Handouts. Data reports used in the study adhered to leading research-based recommendations concerning the best ways in which to present the data in report format, though they did so in a way that did not deviate from what is commonly seen in data systems currently on the market. In other words, reports used in the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study adhered to the better data presentations commonly seen in data systems, but they did not adhere to the best data presentations that – despite being more effective – are not yet commonly seen in student data systems. This was necessary in order to stay true to real world data system environments so results from the study could be generalized to educators' true data system reporting environments (see *Chapter 2: Literature Review: History of Specific Research Contributions* for a historic timeline of research-based recommendations for report design, which relate to framing).

The report sets participants received all contained the same data. For example, all participants were viewing the same data as each other when viewing Report 1, just as they were viewing the same data as each other when viewing Report 2. Thus the data was not "real" data of the participants' own students and school sites. Keeping the data the same was vital for two main reasons:

- The measurement of each participant's data analysis accuracy can be compared to that of the other participants with parity.
- The data in both of the two reports was carefully selected so the most common *incorrect* approaches to analyzing data from each particular assessment on which the data was based did not result in the same answers as the *correct* approaches to analyzing the data. For example, educators analyzing CST

content cluster data often make the mistake of assuming the cluster with the highest score is a site's most likely strength. However, CST content clusters only gain such meaning when compared to state performance such as that of the State Minimally Proficient (SMP) since the clusters differ in difficulty (CDE, 2012). Thus data was used where the cluster in which the site most exceeded the performance of the SMP did not happen to be the same cluster with the highest score. Therefore educators making the most common faulty analyses would not be mistaken for educators making correct analyses, and thus the data would remain as indicative as possible of the *nature* of educators' data analyses: correct versus incorrect.

Data analysis supports used in the study adhered to research-based best practices to the fullest extent possible. For example, small fonts make parts of reports hard to read (Sabbah, 2011) and computer-generated reports for adult audiences should feature at least 2pt spacing between lines; 12pt Times New Roman, Arial, or Tahoma font; and 75-100 characters per line in order to improve the reports' interpretation (Leeson, 2006). Also, over-the-counter medication labels should be at least 1.2mm in vertical height and no more than 40 characters per inch, with appropriate type size include letter contrast, line spacing, print and background color, and type style to increase legibility (Watanabe, Gilbreath, & Sakamoto, 1994). Thus footers provided to assist report analyses conformed to these specifications. However, given the controversies concerning framing, each support was framed in two different ways (see *Chapter 2: Literature Review: History of Specific Research Contributions* for a sampling of research-based recommendations considered when determining the two ways in which the data analysis supports in this

study conformed). In order to mimic real-world conditions, the abstracts and interpretation guides addressed all major questions the reports were designed to answer, as opposed to being geared exclusively toward the questions asked in this study's survey. This way the documentation best mimicked the most effective abstracts and interpretation guides used in real situations, as there are multiple tasks for which educators might be using data reports in real life, and the documentation used in this study could not be unfairly geared only toward the questions asked in the study survey. Please see *Appendix C* for printed copies of the 8½" x 11" handouts respondents received. The following details are summarized in *Table 3.06*.

Table 3.06: *Format of Report 1 and 2 Handouts Distributed to Study Participants*

Folder	Report/Footer	Abstract	Interpretation Guide
White	No Footer (Plain Report)	No Abstract	No Guide
Green	Footer A (Shorter)	No Abstract	No Guide
Yellow	Footer B (Longer)	No Abstract	No Guide
Purple	No Footer (Plain Report)	Abstract A (Less Dense)	No Guide
Blue	No Footer (Plain Report)	Abstract B (Denser)	No Guide
Black	No Footer (Plain Report)	No Abstract	Guide A (2 Pages)
Red	No Footer (Plain Report)	No Abstract	Guide B (3 Pages)

Scenario 1: Control group (white folders). Respondents receiving no added analysis supports received the following handout in their first folder, which they used to answer Questions 4-5:

- Report 1 with No Footer (as labeled in *Appendix C*)

They later received the following handout in their second folder to answer Questions 6-7:

- Report 2 with No Footer (as labeled in *Appendix C*)

When these respondents reached Question 8 they answered its first version.

Scenario 2: Footers in Style A (green folders). Respondents receiving footers on their reports that were shorter and slightly less wordy (1^{st} report footer: 39 words, 186 characters without spaces, 224 characters with spaces; 2^{nd} report footer: 34 words, 156 characters without spaces, 228 characters with spaces) than the alternatively-framed footers and contained headings that utilized text color with meaning received the following handout in their first folder, which they used to answer Questions 4-5:

- Report 1 with Footer A (as labeled in *Appendix C*)

They later received the following handout in their second folder to answer Questions 6-7:

- Report 2 with Footer A (as labeled in *Appendix C*)

When these respondents reached Question 8 they answered its second version.

Scenario 3: Footers in Style B (yellow folders). Respondents receiving footers on their reports that were longer and slightly wordier (1^{st} report footer: 58 words, 269 characters without spaces, 324 characters with spaces; 2^{nd} report footer: 42 words, 199 characters without spaces, 237 characters with spaces) than the alternatively-framed footers and contained no headings or colored text received the following handout in their first folder, which they used to answer Questions 4-5:

- Report 1 with Footer B (as labeled in *Appendix C*)

They later received the following handout in their second folder to answer Questions 6-7:

- Report 2 with Footer B (as labeled in *Appendix C*)

When these respondents reached Question 8 they answered its second version.

Scenario 4: Abstracts in Style A (purple folders). Respondents whose reports were accompanied by abstracts that were less dense and contained less information than the alternatively-framed abstracts and utilized heading color with meaning received the following handouts in their first folder, which they used to answer Questions 4-5:

- Report 1 with No Footer (as labeled in *Appendix C*)
- Report 1 Abstract A (as labeled in *Appendix C*)

They later received the following handouts in their second folder to answer Questions 6-7:

- Report 2 with No Footer (as labeled in *Appendix C*)
- Report 2 Abstract A (as labeled in *Appendix C*)

When these respondents reached Question 8 they answered its third version.

Scenario 5: Abstracts in Style B (blue folders). Respondents whose reports were accompanied by abstracts that were more dense and contained more information than the alternatively-framed abstracts and did not utilize heading color with meaning received the following handouts in their first folder, which they used to answer Questions 4-5:

- Report 1 with No Footer (as labeled in *Appendix C*)
- Report 1 Abstract B (as labeled in *Appendix C*)

They later received the following handouts in their second folder to answer Questions 6-7:

- Report 2 with No Footer (as labeled in *Appendix C*)
- Report 2 Abstract B (as labeled in *Appendix C*)

When these respondents reached Question 8 they answered its third version.

Scenario 6: Interpretation guides in Style A (black folders). Respondents whose reports were accompanied by interpretation guides that were shorter and contained less information (two pages) than the alternatively-framed guides (three pages) and utilized heading color with meaning received the following handouts in their first folder, which they used to answer Questions 4-5:

- Report 1 with No Footer (as labeled in *Appendix C*)
- Report 1 Interpretation Guide A (as labeled in *Appendix C*)

They later received the following handouts in their second folder to answer Questions 6-7:

- Report 2 with No Footer (as labeled in *Appendix C*)
- Report 2 Interpretation Guide A (as labeled in *Appendix C*)

When these respondents reached Question 8 they answered its fourth version.

Scenario 7: Interpretation guides in Style B (red folders). Respondents whose reports were accompanied by interpretation guides that were longer and contained more information (three pages) than the alternatively-framed guides (two pages) and did not utilize heading color with meaning received the following handouts in their first folder, which they used to answer Questions 4-5:

- Report 1 with No Footer (as labeled in *Appendix C*)
- Report 1 Interpretation Guide B (as labeled in *Appendix C*)

They later received the following handouts in their second folder to answer Questions 6-7:

- Report 2 with No Footer (as labeled in *Appendix C*)
- Report 2 Interpretation Guide B (as labeled in *Appendix C*)

When these respondents reached Question 8 they answered its fourth version.

Triangulation. While this was a quantitative rather than a mixed-methods study, there were still opportunities for triangulation. Although one sampling strategy was utilized, collecting data from a variety of educators leant data triangulation to the study. Also, each report respondents analyzed in the study were used for two different data analysis questions rather than one, and two reports were used in this way so as to provide a total of four data analysis questions. Report differences, to which all 211 participants were exposed, included:

- Report 1 was graphical in format, whereas Report 2 was tabular in format.
- Report 1 utilized the use of a key/legend to answer analysis questions, whereas Report 2 did not.
- Color was vital to the understanding of Report 1 data, whereas color was not pertinent to the analysis of Report 2 data.
- Report 1 related to an assessment considered higher stakes than the Report 2 assessment.
- Report 1 presented aggregate data in the form of site and state averages, whereas Report 2 presented student-level data.

Question differences, to which all 211 participants were exposed, included:

- Questions 4-5 analyses required more steps than Questions 6-7 analyses, presenting varied levels of critical thinking and difficulty.
- Questions 4-5 each required the selection of only one of the multiple-choice answer options, whereas Questions 6-7 each required the selection of one or more of the multiple-choice answer options, with the correct number of selections that must be made left as undefined for respondents as the correct answers.

These variations leant within-method methodological triangulation to the study.

Questionnaire coding. A code book was created prior to administering the survey. While code descriptions are featured below, see A*ppendix D* for the code book, featuring details on the coding process used on the data file. Coding was assisted by the use of Google Docs. For example, the coding process requires adding an identification number to the first field of each questionnaire – and thus each respondent's record – on a data spreadsheet, adding category headers to the top of each data column, and organizing responses as one person/questionnaire per row (Tufféry, 2011). Google Docs Form accomplished all of these aspects automatically each time any respondent submitted his or her questionnaire.

Questions 1-2 were coded as follows because the answer options were likely to be more common nearest to option "a," as this was the least demanding answer option, and less common as they neared option "e," as this was the most demanding answer option:

1. How long have you worked as an educator (e.g., teacher or administrator) for students under 19 years of age? *Select the highest option applicable.*
    a. less than 1 year [*assign 1 point*]
    b. 5 years [*assign 2 points*]

c. 10 years [*assign 3 points*]

d. 15 years [*assign 4 points*]

e. 20 or more years [*assign 5 points*]

2. Which of the following roles best describes your current position? *If your role is mixed, select the role requiring most of your time.*

a. Teacher [*assign 1 point*]

b. Colleague Coach (e.g., Teacher on Special Assignment) [*assign 2 points*]

c. Site/School Administrator [*assign 3 points*]

d. District Administrator [*assign 4 points*]

The point order used for Question 1-2 was reversed for Question 3. This suited the practice of assigning low or negative numbers for disagreeing/negative answer responses, as the answer options were likely to be more positive nearest to option "a" and more negative as they neared option "d."

3. How proficient are you at analyzing student performance data? *In your opinion:*

a. Very proficient [*assign 4 points*]

b. Somewhat proficient [*assign 3 points*]

c. Not proficient [*assign 2 points*]

d. Far from proficient [*assign 1 point*]

For Questions 4-7, zero points were assigned to answers that were incorrect, and 1 point to answers that were correct, with only one answer being accepted per question. This was possible because each of these questions only had one clearly correct answer and assessed responders' ability to correctly analyze the data they had been given via the particular report they had been given:

4. Which content cluster is most likely the School's strength? *Base your answer on the folder's Report 1.* [*assign 1 point for correct answer, and 0 points for any incorrect answer*]
    a. Word Analysis and Vocabulary Development [*assign 0 points*]
    b. Reading Comprehension [*assign 0 points*]
    c. Literary Response and Analysis [*assign 0 points*]
    d. Written Conventions [*assign 0 points*]
    e. Writing Strategies [*assign 1 point*]
    f. Writing Applications [*assign 0 points*]
5. Which content cluster is most likely the School's weakness? *Base your answer on the folder's Report 1.* [*assign 1 point for correct answer, and 0 points for any incorrect answer*]
    a. Word Analysis and Vocabulary Development [*assign 0 points*]
    b. Reading Comprehension [*assign 0 points*]
    c. Literary Response and Analysis [*assign 0 points*]
    d. Written Conventions [*assign 1 point*]
    e. Writing Strategies [*assign 0 points*]
    f. Writing Applications [*assign 0 points*]
6. Which student(s) did NOT score Proficient on the CELDT? Check all that apply. *Base your answer on the folder's Report 2. CHECK ALL THAT APPLY.* [*assign 1 point for correct answer, and 0 points for any incorrect answer*]
    a. Student B and Student D [*assign 1 point*]
    b. *Any Other Answer or Answer Combination* [*assign 0 points*]

7. In which area(s) did at least 1 student earn a score that PREVENTED him/her from scoring Proficient on the CELDT? *Base your answer on the folder's Report 2. CHECK ALL THAT APPLY.* [*assign 1 point for correct answer, and 0 points for any incorrect answer*]
    a. Speaking and Overall [*assign 1 point*]
    b. *Any Other Answer or Answer Combination* [*assign 0 points*]

Not Numbered. What color is your folder? *The cover of your report materials folder features the name of its color.* [folders are also colored in entirety to match their color names] [*assign points based on increasing levels/volume of text support*]

a. White [*assign 1 point*]
b. Yellow [*assign 3 points*]
c. Green [*assign 2 points*]
d. Blue [*assign 5 points*]
e. Purple [*assign 4 points*]
f. Red [*assign 7 points*]
g. Black [*assign 6 points*]

Question 8 varied based on the data analysis support each respondent received. The point order used for Question 8 followed the practice of assigning low or negative numbers for disagreeing/negative answer responses, as the answer options were likely to be more positive nearest to option “a” and more negative as they neared option “d.” The four different versions of Question 8 are listed below as Question 8a for respondents with no data analysis supports, Question 8b for respondents with footers, Question 8c for respondents with abstracts, and Question 8d for respondents with interpretation guides:

8a. The 2 reports you just used did not offer any special assistance in analyzing the data. If they had been accompanied by text (e.g., a footer, guide, or abstract) designed to help you interpret the data, would you likely have used the added support?

a. Yes – I probably would use the support. [*assign 4 points for identification but convert to 0 for analyses involving whether or not support was present or used*]
b. No – I probably would not use the support. [*assign 1 point for identification but convert to 0 for analyses involving whether or not support was present or used*]

8b. The 2 reports you just used contained footers with analysis guidelines designed to help you. Did you read these footers *before* answering questions related to the reports?

a. Yes – I referred to both reports' footers. [*assign 4 points for identification but convert to 1 for analyses involving whether or not support was used for Questions 4-7*]
b. I referred to Report 1's footer but not Report 2's footer. [*assign 3 points for identification but convert to 1 for analyses involving whether or not support was used for Questions 4 and 5 and convert to 0 for analyses involving whether or not support was used for Questions 6 and 7*]
c. I referred to Report 2's footer but not Report 1's footer. [*assign 2 points for identification but convert to 0 for analyses involving whether or not support was used for Questions 4 and 5 and convert to 1 for analyses involving whether or not support was used for Questions 6 and 7*]

d. No – I did not refer to either footer. [*assign 1 point but convert to 0 for analyses involving whether or not support was used for Questions 4-7*]

8c. The 2 reports you just used were each accompanied by a 1-page abstract (like a reference sheet) with analysis guidelines designed to help you. Did you read these abstracts/sheets before answering questions related to the reports?

a. Yes – I referred to both reports' abstracts/sheets. [*assign 4 points but convert to 1 for analyses involving whether or not support was used for Questions 4-7*]

b. I referred to Report 1's abstract/sheet but not Report 2's abstract/sheet. [*assign 3 points for identification but convert to 1 for analyses involving whether or not support was used for Questions 4 and 5 and convert to 0 for analyses involving whether or not support was used for Questions 6 and 7*]

c. I referred to Report 2's abstract/sheet but not Report 1's abstract/sheet. [*assign 2 points for identification but convert to 0 for analyses involving whether or not support was used for Questions 4 and 5 and convert to 1 for analyses involving whether or not support was used for Questions 6 and 7*]

d. No – I did not refer to either abstract/sheet. [*assign 1 point but convert to 0 for analyses involving whether or not support was used for Questions 4-7*]

8d. The 2 reports you just used were each accompanied by an interpretation guide (a packet) with analysis guidelines designed to help you. Did you read these guides before answering questions related to the reports?

a. Yes – I referred to both reports' guides. [*assign 4 points but convert to 1 for analyses involving whether or not support was used for Questions 4-7*]

b. I referred to Report 1's guide but not Report 2's guide. [*assign 3 points for identification but convert to 1 for analyses involving whether or not support was used for Questions 4 and 5 and convert to 0 for analyses involving whether or not support was used for Questions 6 and 7*]

c. I referred to Report 2's guide but not Report 1's guide. [*assign 2 points for identification but convert to 0 for analyses involving whether or not support was used for Questions 4 and 5 and convert to 1 for analyses involving whether or not support was used for Questions 6 and 7*]

d. No – I did not refer to either guide. [*assign 1 point but convert to 0 for analyses involving whether or not support was used for Questions 4-7*]

Questions 9-10 were coded as follows because the answer options were likely to be more common nearest to option "a," as this was the least demanding answer option, and less common as they neared option "e," as this was the most demanding answer option:

9. Lots of professional development happens at school sites: for example, demonstrations to accompany textbook adoptions, meetings with colleagues to share differentiation strategies, training on how to use new software, etc. Only some professional development specifically focuses on how to analyze student data. Within the last 12 months, how many hours of professional development have you had that specifically focused on teaching you how to correctly interpret student data*? Select the highest option applicable. Time spent*

analyzing student data without guidance should not be counted, nor should time spent learning technology to generate student data.

a. 0 hours [*assign 1 point*]

b. 1 hour [*assign 2 points*]

c. 2 hours [*assign 3 points*]

d. 5 hours [*assign 4 points*]

e. 8 or more [*assign 5 points*]

10. *Educational Measurement* refers to the analysis of student assessment data to draw conclusions about abilities. How many graduate-level courses have you taken that were *specifically dedicated* to educational measurement (e.g., student performance data analysis, measurement theory, or psychometrics)? *Select the highest option applicable.*

a. 0 courses [*assign 1 point*]

b. 1 course [*assign 2 points*]

c. 2 courses [*assign 3 points*]

d. 3 courses [*assign 4 points*]

e. 4 or more [*assign 5 points*]

After the coding of all 211 rows of respondent data and in each of the Columns A and Q-JH noted in the code book in *Appendix D*, four rows were added to the bottom of the data file and were filled with formulas to make the following calculations of the same-column cells within the range of the 211 respondent data rows:

- sum/total the cell contents/values
- count non-blank cells

- divide the above count of non-blank cells by 211 to calculate % of participants
- mean/average the cell contents/values

The values in the four rows described above were used in *Tables 4.01-4.14*, which are each disaggregated by reporting environment or demographics and give mean values for the impact support presence and support use has on educators' data analysis accuracy.

Operational Definitions of Variables

Please see *Appendix B* for the actual Google Docs Form survey that was used for the study. *Table 3.07* illustrates the survey's variables, as well as the corresponding questions and scales designed to address them. Question verbiage and the exact scale value attributed to each answer option are featured in the previous section (see *Chapter 3: Research Method: Materials/Instruments: Questionnaire coding*).

Table 3.07: *Survey Variables, Research Questions, Survey Items, & Scales*

Variable Name	Research Questions	Survey Item(s) and Scale
Independent Variable 1: Support Provided & Framing Style	Descriptive Question 1: Which analysis support did the educator receive? Related to Research Questions: Q1, Q2a, Q2b, Q3a, Q3b, Q4a, Q4b	See unnumbered Survey Question 7b: support provided (ordinal scale)

Independent Variable 2: Use of Analysis Support	Descriptive Question 2: To what extent did respondents use added analysis support? Related to Research Questions: Q1, Q2a, Q2b, Q3a, Q3b, Q4a, Q4b, Q5a-Q5e, Q6a-Q6e	See Survey Question 8; support usage (ordinal scale)
Independent Variable 3: School Site Level Type	Descriptive Question 3: What was the educators' school site level type? *Note: This variable was included because school level type is sometimes rumored to have an impact on data analysis practice and competency.* Related to Research Question: Q5a	Not on the survey (based on public California Department of Education data added to Column S on the data file based on each participant's school site, and coded in Column JG) (ordinal scale)
Independent Variable 4: School Site Level	Descriptive Question 4: What was the educators' school site level? *Note: This variable was included because school level is sometimes rumored to have an impact on data analysis practice and competency.* Related to Research Question: Q5b	Not on the survey (based on public California Department of Education data added to Column S on the data file based on each participant's school site, and coded in Column JG) (ordinal scale)

Independent Variable 5: Academic Performance	Descriptive Question 5: What was the Growth API of the school site? *Note: This variable was included because teachers of students with more significant struggles are often rumored to have more data analysis practice and competency.* Related to Research Question: Q5c	Not on the survey (based on public California Department of Education data added to Column W on the data file based on each participant's school site) (ordinal scale)
Independent Variable 6: English Learner Population	Descriptive Question 6: What was the population of English Learners attending the school site? *Note: This variable was included because teachers of students with more significant struggles are often rumored to have more data analysis practice and competency.* Related to Research Question: Q5d	Not on the survey (based on public California Department of Education data added to Column X on the data file based on each participant's school site) (ordinal scale)

Independent Variable 7: Socioecon-omically Disadvantaged Population	Descriptive Question 7: What was the population of Socioeconomically Disadvantaged students attending the school site? *Note: This variable was included because teachers of students with more significant struggles are often rumored to have more data analysis practice and competency.* Related to Research Question: Q5e	Not on the survey (based on public California Department of Education data added to Column Y on the data file based on each participant's school site) (ordinal scale)
Independent Variable 8: Students with Disabilities Population	Descriptive Question 8: What was the population of Students with Disabilities attending the school site? *Note: This variable was included because teachers of students with more significant struggles are often rumored to have more data analysis practice and competency.* Related to Research Question: Q5f	Not on the survey (based on public California Department of Education data added to Column Z on the data file based on each participant's school site) (ordinal scale)

Independent Variable 9: Veteran Status	Descriptive Question 9: How long had the educator been working as an educator? Related to Research Question: Q6a	See Survey Question 1: years teaching (ordinal scale)
Independent Variable 10: Role	Descriptive Question 10: What was the educator’s current role? Related to Research Question: Q6b	See Survey Question 2: job title (ordinal scale)
Independent Variable 11: Perceived Proficiency	Descriptive Question 11: What was the educator’s perceived level of data analysis proficiency? Related to Research Question: Q6c	See Survey Question 3: perceived data analysis proficiency (ordinal scale)
Independent Variable 12: Professional Development	Descriptive Question 12: Within the last year, how many hours of training/professional development had the educator participated in that specifically focused on how to properly analyze student data? Related to Research Question: Q6d	See Survey Questions 9: hours of related training/professional development (ordinal scale)

Independent Variable 13: Graduate-Level Course Instruction	Descriptive Question 13: How many graduate-level educational measurement courses had the educator taken? Related to Research Question: Q6e	See Survey Questions 10: number of related graduate-level course instruction (ordinal scale)
Dependent Variable 1: Data Analysis Accuracy	Descriptive Question 14: How accurate was the educator's analysis of student achievement data? Related to Research Questions: Q1, Q2a, Q2b, Q3a, Q3b, Q4a, Q4b, Q5a-Q5e, Q6a-Q6e	See Survey Questions 4-7: content cluster strength, content cluster weakness, strongest grade-level performance, weakest grade-level performance (nominal scale)

Behavioral economics' impact on variables. Regarding behavioral economics' impact on variables, it is important to note that priming and framing are the most relevant behavioral economics dimensions in this study. The complete process of data-informed decision-making is influenced by behavioral economics facets such as priming, biases, heuristics, prototypes, judgments, anchoring, and framing (Kahneman, 2011). For example, even seemingly insignificant differences in how content is arranged can have a major impact on the decisions people make based on that content (Thaler & Sunstein, 2008). However, this study was concerned only with the environment of the online data system, as the study's purpose lay in finding ways data systems can be improved to

facilitate improved data analyses. Thus conditions outside of those that can be controlled within a data system were not manipulated or used as variables in the study. Likewise, this study related to improving the accuracy of educators' data analyses in the thought portion – or "data-informed" portion – of data-informed decision-making, as opposed to evaluating the decisions to which the data-informed thoughts lead. Thus study variables were exclusively concerned with behavioral economics dimensions that can be impacted by the data system during data analyses: the priming and framing dimensions.

Priming. Priming is the behavioral economics dimension that a data system can facilitate before data reports are viewed, as this study sought to measure the value of data system analysis supports that can be – though are not always – viewed before educators view the data reports to inform their decisions. Essentially, the analysis support resources can prime the educator's thoughts concerning the data, and then those thoughts can prime the educator's decisions. For example, Goodman, and Hambleton (2004) noted value gained in states that accompanied data reports with information for parents to read before reading and interpreting the data.

Study participants received different data analysis supports (or none) with the potential to prime his or her thoughts and analyses concerning the data reports that were also viewed. However, it was possible that respondents receiving resources with the potential to prime would not use the resources. While *Dependent Variable 1: Data Analysis Accuracy* (Survey Questions 4-7) measured the accuracy of the educator's analyses of the data, *Independent Variable 2:Use of Analysis Support* (Survey Question 8) measured the extent to which the respondent used the added resources. This

combination of question intent allowed the resources' impact on priming to be better determined.

Framing. Framing applies to the presentation of information, and framing the same information to someone in different ways will often result in different levels of difficulty in understanding or analyzing the information (Kahneman, 2003, 2011). The manner in which content is organized for people using it to make decisions significantly impacts those decisions (Thaler & Sunstein, 2008). Framing thus plays a large role in data analysis accuracy and data-informed decision-making. Reports used in this study therefore subscribed to leading research-based recommendations concerning the best ways in which to frame the data in report format, though they did so in a way that did not deviate from what is commonly seen in data systems currently on the market. In other words, reports used in the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study adhered to the better data presentations commonly seen in data systems, but they did not adhere to the best data presentations that – despite being more effective – are not yet commonly seen in student data systems (see *Chapter 2: Literature Review: History of Specific Research Contributions* for a historic timeline of research-based recommendations for report design, which relate to framing).

Suggested ways to present analysis guidance in footers, abstracts, and interpretation guides were utilized in this study, but the best manner in which to frame these resources had not yet been determined in regards to direct impact on analysis accuracy. Thus each of the three support resources used in this study were framed in two different formats for respondents.

Framing was also a key reason behind the necessity of this study. While research already existed concerning the best *data displays* to use to improve educators' analyses, such as using bar graphs rather than pie charts, the best way in which to frame *analysis support* within a data system to improve educators' analyses had not yet been determined. This study was meant to help fill the gap in research literature. Thus the footers, abstracts, and interpretation guides utilized in this study were each presented in two different formats. *Dependent Variable 1: Data Analysis Accuracy* (Survey Questions 4-7) then measured the accuracy of the educator's analysis of the data when exposed to each data system support. This allowed the study to not only highlight the extent to which each data system analysis support can potentially increase analysis accuracy, but also the best manner in which to frame these resources in regards to direct impact on analysis accuracy.

Data Collection, Processing, and Analysis

Data collection method. This experimental study, which was conducted in a computer laboratory environment rather than a field test environment, involved a web-based questionnaire crafted and administered in Google Docs, taking advantage of the Google Docs Form feature, and involved groups of no more than 30 respondents at each administration time. The researcher explored three data analysis supports provided by a data system, each framed in two different formats, by presenting 211 elementary and secondary educators with different versions of the same two student achievement data report environments. These report sets fit into one of the following treatment categories (a) no added analysis support; (b) analysis support by way of footers directly on the reports, which were offered in two different framing styles; (c) analysis support by way

of abstracts, which accompanied the reports and were offered in two different framing styles; or (d) by way of interpretation guides, which accompanied the reports and were offered in two different framing styles (see *Appendix C* for reports and handouts).

While an online survey was used, the interviewer was present and available to participants during the survey completion process to compensate for the online format's weakness of otherwise not allowing for participants to ask questions and receive clarification if the survey process confuses them. Such answers and clarifications were restricted to information to help participants in survey completion but not in relation to any matters that could bias the results. For example, participants received reports meant to be used in answering particular data analysis questions. If a respondent asked, "Am I using this report and the previous report to answer this question?" it would be acceptable to answer, "No; you will only use the second report." However, if the respondent asked, "Which columns on this report's table should I be looking at?" the interviewer had to respond, "I'm sorry, but I cannot answer that question." The introduction prior to the survey addressed the fact that analysis questions would be inappropriate for the interviewer to answer.

The efficiency and cost-effectiveness of this electronic approach to data collection allowed for a larger number of participants, which likely resulted in a more reliable data sampling than would be possible with smaller numbers. The larger number also better allowed for a thoroughly cross-sectional sampling. In addition, this method leant itself well to the editing and coding phases of the study.

Sampling and materials. Please see the *Participants* section of this chapter for sampling procedures and details, which relate to data collection and analysis. Please see

the *Materials/Instruments* section of this chapter for details on the survey and handouts used, as well as ways in which they facilitate triangulation, all of which relate to data collection, processing, and analysis.

Priori power analysis. Each participant's survey translated into a percent correct score for Questions 4-7. Their accuracy was thus reported much like the U.S. Department of Education reports accuracy in relation to its studies shadowing NCLB, such as that of USDEOPEPD (2010). To determine ideal sample size through priori power analysis, the researcher conducted a two-tailed t-test calculating the difference between two independent means utilizing the G*Power 3.1 statistical analysis tool. For this analysis's details, see *Figure 3.01* for all input and output parameters and see *Figure 3.02* for an X-Y plot graph showing the power (1- β probability of correctly rejecting the null hypothesis) in relation to sample size. Input parameters included: tails = two, effect size d = 0.5, α error of probability (alpha, the probability of a type I error) = 0.05, and power (1-β error of probability for a type II error) = 0.95. Output parameters included noncentrality parameter δ = 3.6228442, critical t = 1.9714347, Df = 208, sample size group1 = 105, sample size group 2 = 105, total sample size, = 210, actual power = 0.9501287. The priori two-tailed t-test thus resulted in a recommended sample size of at least 210 educators.

However, the researcher also conducted an F-test linear multiple regression analysis, fixed model, R^2 deviation from zero, using the G*Power 3.1 statistical analysis tool. For this analysis's details, see *Figure 3.03* for all input and output parameters and see *Figure 3.04* for an X-Y plot graph showing the power (1- β probability of correctly rejecting the null hypothesis) in relation to sample size. Input parameters included: effect

size $f^2 = 0.15$, α error of probability (alpha, the probability of a type I error) = 0.05, power (1-β error of probability for a type II error) = 0.95, and number of predictors based on independent variables =7. Output parameters included noncentrality parameter λ = 22.9500000, critical F = 2.0732820, numerator df = 7, denominator df = 145, total sample size = 153, and actual power = 0.9503254. The priori F-test thus resulted in a recommended sample size of at least 153 educators. However, since the 210 sample size resulting from the two-tailed t-test was greater than 153, a sample size of at least 210 educators was used as the goal for this study. 211 participants were thus ultimately involved.

Editing the data. The editing phase typically involves editing for omissions, legibility, and preparing the data for storage and coding (Tufféry, 2011). The use of Google Docs Form for this study helped to handle these aspects. For example, the Google Docs "required question" setting was assigned to each survey question to eliminate the risk of omissions. This approach simultaneously eliminated the risk of response bias resulting from nonresponses on the survey. However, descriptive information describing respondents and non-respondents would have been used in the event that some participants did not complete their surveys, but this was not necessary since all participants finished the survey. Non-respondent data on the school sites where participants were working came from the California Department of Education's DataQuest site (http://data1.cde.ca.gov/dataquest), where the researcher generated a 2011-12 (the most recent school year available) School Quality Snapshot report for each site to determine its:

- 2012 Growth Academic Performance Index (API), California's state accountability measure and also a factor in Adequate Yearly Progress (AYP) for California's federal accountability
- English Learner (EL) population
- Socioeconomically Disadvantaged population
- Students with Disabilities population

Likewise, legibility issues were not problematic, as respondents entered their answers electronically. This simultaneously handled readying the data for storage, as it was automatically stored securely and password-protected online as soon as it was entered, and it made strides in preparing the data coding (addressed below).

Table 3.08: *Linear Regression Analyses Applied to Research Question Variables*

Abbreviated Research Question	Relationship	Explanation of Relationship
Q1. Support's impact on analysis accuracy	A = f(S)	A (Analysis Accuracy) is a function of S (Support)
Q2a. Footer's impact on analysis accuracy	A = f(F)	A (Analysis Accuracy) is a function of F (Footer)
Q2b. Footer framing's impact on analysis accuracy	A = f(FF)	A (Analysis Accuracy) is a function of F (Footer's Framing)
Q3a. Abstract's impact on analysis accuracy	A = f(B)	A (Analysis Accuracy) is a function of B (Abstract)

Q3b. Abstract framing's impact on analysis accuracy	A = f(BF)	A (Analysis Accuracy) is a function of BF (Abstract's Framing)
Q4a. Interpretation guide's impact on analysis accuracy	A = f(I)	A (Analysis Accuracy) is a function of I (Interpretation Guide)
Q4b. Interpretation guide framing's impact on analysis accuracy	A = f(IF)	A (Analysis Accuracy) is a function of IF (Interpretation Guide's Framing)

Regression analysis. *Table 3.08* illustrates the survey's research question variables with the corresponding regression analysis features designed to address them. Results from the study were downloaded into a Microsoft Excel® worksheet and coded according to *Chapter 3: Research Method: Materials/Instruments: Questionnaire coding* and *Appendix D*. A scatterplot (see *Figure 3.05*) was then fashioned in Microsoft Excel® to show the distribution of respondent's data analysis accuracy scores (0%-100%) in relation to their reporting environments (1-7). A linear regression trend line displaying the trend line equation of $y = 0.0003x + 0.232$ and the R^2 value of $R^2 = 0.0016$ was added to the scatterplot. This figure, along with other tools yet to be discussed, was used in making casual observations concerning the supports' impact on respondents' analysis accuracy. However, *Figure 3.05* was not used to determine the degree to which each support impacted data analysis accuracy, as a scatterplot is not designed to render such determinations.

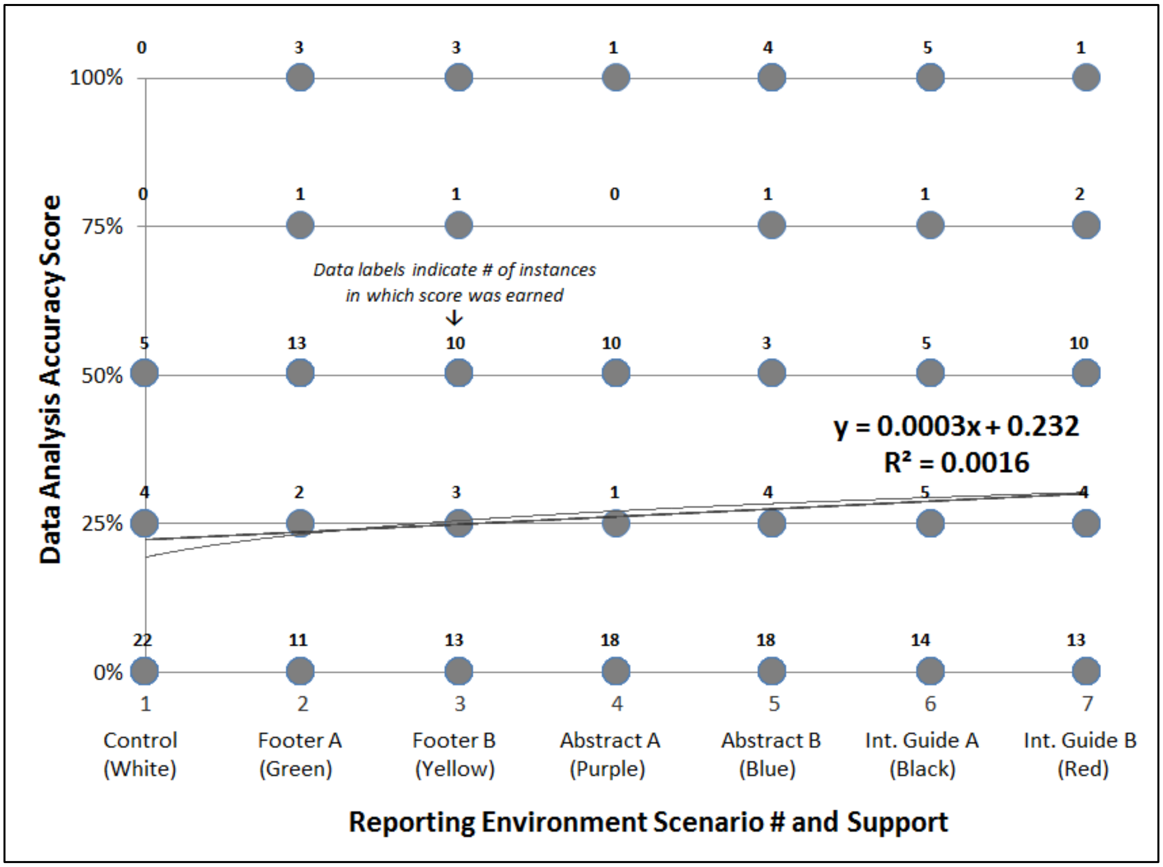


Figure 3.05: *Distribution of Data Analysis Accuracy Scores with Multiple Points Overlaid*

One of multiple graphical displays of data (*Figure 3.05*) was initially used for casual observations concerning the supports' impact on respondents' analysis accuracy. However, this format was not designed to facilitate analyses of the degree to which each support impacted data analysis accuracy. Mathematical equations expressing functional relationships was thus needed (Schroeder, Sjoquist, & Stephan, 1986). Assuming the null hypothesis that each support has no significant impact on data analysis accuracy, the form of the equations was straight lines. *Table 3.09* illustrates these functional relationship formulas.

Table 3.09: *Linear Regression Relationship Applied to Research Question Variables*

Abbreviated Research Question	Relationship	Explanation of Relationship
Q1. Support's impact on analysis accuracy	$A = \alpha+\beta S$	A (Analysis Accuracy) with unknown parameters (α and β) holding for the education population is a function of S (Support) with unknown parameters (α and β) holding for the education population
Q2a. Footer's impact on analysis accuracy	$A = \alpha+\beta F$	A (Analysis Accuracy) with unknown parameters (α and β) holding for the education population is a function of F (Footer) with unknown parameters (α and β) holding for the education population
Q2b. Footer framing's impact on analysis accuracy	$A = \alpha+\beta FF$	A (Analysis Accuracy) with unknown parameters (α and β) holding for the education population is a function of F (Footer's Framing) with unknown parameters (α and β) holding for the education population

Q3a. Abstract's impact on analysis accuracy	A = α+βB	A (Analysis Accuracy) with unknown parameters (α and β) holding for the education population is a function of B (Abstract) with unknown parameters (α and β) holding for the education population
Q3b. Abstract framing's impact on analysis accuracy	A = α+βBF	A (Analysis Accuracy) with unknown parameters (α and β) holding for the education population is a function of BF (Abstract's Framing) with unknown parameters (α and β) holding for the education population
Q4a. Interpretation guide's impact on analysis accuracy	A = α+βI	A (Analysis Accuracy) with unknown parameters (α and β) holding for the education population is a function of I (Interpretation Guide) with unknown parameters (α and β) holding for the education population

Q4b. Interpretation guide framing's impact on analysis accuracy	$A = \alpha + \beta IF$	A (Analysis Accuracy) with unknown parameters (α and β) holding for the education population is a function of IF (Interpretation Guide's Framing) with unknown parameters (α and β) holding for the education population

The researcher set the values of these equations' population parameters using the sample of 211 educators that was used for this study. Essentially, the researcher determined whether the slope (β) was greater than zero to estimate whether use of the given support resulted in an increase in data analysis accuracy, and the researcher estimated the value of β to estimate the extent of the support's impact on data analysis accuracy. To find a linear approximation of variable relationships, the researcher used the Microsoft Excel® linear trend/regression line function to insert a straight line between the points on each scatterplot (as seen in *Figure 3.05*). This was important for even casual observations, as it accounted for the fact that the scatterplot featured identical responses with a single mark; for example, 22 of the 31 people receiving no analysis support received an analysis score of 0%, whereas four people receiving no analysis support received an analysis score of 25%, yet the two analysis scores were merely represented by two equally-sized marks on a scatterplot. However, the researcher added the number of instances above each data point for added clarification. The researcher also opted to display the regression line's equation and the R-squared value on the scatterplot (as seen in *Figure 3.05*).

Independent Samples T-Tests. The researcher used SPSS for the Independent Samples T-Tests for the analyses of nominal and scale data. These were used to investigate the relationship between data analysis support *use* and data analysis accuracy. This test compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who received the support and those who did not). Four such tests were conducted in order to examine the impact of four different types of support use: (a) any support, combining the supports that follow as b-d; (b) footer; (c) abstract; and (d) interpretation guide. Respondent data rows were sorted by the contents of Column Q to sort responses by reporting environment to facilitate upcoming analyses. The outputs for these SPSS Independent Samples T-Tests are featured in *Appendices E-H*.

The researcher conducted Independent Samples T-Tests to investigate the relationship between data analysis support *presence* and data analysis accuracy. Four such tests were conducted in order to examine the impact of the *presence* of four different types of supports: (a) any support, combining the supports that follow as b-d; (b) footer; (c) abstract; and (d) interpretation guide. These investigations concerning support *presence* differed from the investigations concerning support *use* in that the former was concerned merely with whether or not a support was present and did not concern whether or not the respondent used any support. Conversely, the latter was concerned with whether or not the respondent indicated he or she actually used a support. Respondent data rows were sorted by the contents of Column Q to sort responses by reporting environment to facilitate upcoming analyses. The outputs for these SPSS Independent Samples T-Tests are featured in *Appendices I-L*.

The researcher also conducted Independent Samples T-Tests to determine if each embedded data analysis support's format had a significant impact on its effectiveness. This test leant itself well to this investigation, as two different formats were used for each support in the study, creating three pairs to be investigated separately. Three such tests were conducted in order to examine the format set for each of the three different types of supports used in the study: (a) footer; (b) abstract; and (c) interpretation guide. Respondent data rows remained sorted by the contents of Column Q to sort responses by reporting environment to facilitate upcoming analyses. The outputs for these SPSS Independent Samples T-Tests are featured in *Appendices M-O*.

***Support use and data analysis accuracy*.** The researcher needed to determine the relationship between whether or not any data analysis support (e.g., footer, abstract, or interpretation guide) was used by the respondent, as indicated by the respondent, and the resultant data analysis accuracy. The respondent data row contents of Columns IQ and IR were added to SPSS, followed (underneath) by that of Columns IS and IT, then IU and IV, and then IW and IX (see *Appendix D* for column code book descriptions). This created two columns and 844 rows of respondent data in order to include data for all four data analysis questions that were answered by each respondent, each with a chance of a data analysis support being used or not used. For example, a participant might have used the support for Report 1 and obtained a data accuracy score of 100% on related Questions 4 and 5, but then might not have used the support for Report 2 and obtained a data accuracy score of 50% on related Questions 6 and 7. The accuracy rate had to be tied directly to whether or not the support was used and thus tied to each reporting/support instance. Variable settings used for this data are is shown in F*igure 3.06*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	8	0	Analysis Accuracy (% Correct)	None	None	8	Left	Scale	Target
SupportUse	Numeric	8	0	Support Use (0 Not Used, 1 Used)	None	None	8	Left	Nominal	Input

Figure 3.06: *Support Use and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and support use as the grouping variable. This resulted in the statistics shown in *Appendix E*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between support use and data analysis accuracy was significance.

***Footer use and data analysis accuracy*.** The researcher needed to determine the relationship between whether or not a footer was used by the respondent, as indicated by the respondent, and the resultant data analysis accuracy. In order to isolate only instances where a footer was used or not used, which included data from control group participants and participants who were given a reporting environment with footers, contents were *only* used from rows where the value of Column Q equaled 1, 2, or 3 (see *Appendix D* for column code book descriptions). The applicable respondent data row contents of Columns IQ and IR were added to SPSS, followed (underneath) by that of Columns IS and IT, then IU and IV, and then IW and IX. This created two columns and 364 rows of applicable respondent data in order to include data for all four data analysis questions that were answered by each respondent, each with a chance of a footer being used or not used. For example, a participant might have used the footer for Report 1 and obtained a data accuracy score of 100% on related Questions 4 and 5, but then might not have used the footer for Report 2 and obtained a data accuracy score of 50% on related Questions 6 and

7. The accuracy rate had to be tied directly to whether or not the footer was used and thus tied to each reporting/support instance. Variable settings used for this data are is shown in F*igure 3.07*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	8	0	Analysis Accuracy (% Correct)	None	None	8	Left	Scale	Target
FooterUse	Numeric	8	0	Footer Use (Not Used, 1 Used)	None	None	8	Left	Nominal	Input

Figure 3.07: *Footer Use and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and footer use as the grouping variable. This resulted in the statistics shown in *Appendix F*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between footer use and data analysis accuracy was significance.

***Abstract use and data analysis accuracy*.** The researcher needed to determine the relationship between whether or not an abstract was used by the respondent, as indicated by the respondent, and the resultant data analysis accuracy. In order to isolate only instances where an abstract was used or not used, which included data from control group participants and participants who were given a reporting environment with abstracts, contents were *only* used from rows where the value of Column Q equaled 1, 4, or 5 (see *Appendix D* for column code book descriptions). The applicable respondent data row contents of Columns IQ and IR were added to SPSS, followed (underneath) by that of Columns IS and IT, then IU and IV, and then IW and IX. This created two columns and 364 rows of applicable respondent data in order to include data for all four data analysis questions that were answered by each respondent, each with a chance of an abstract being

used or not used. For example, a participant might have used the abstract for Report 1 and obtained a data accuracy score of 100% on related Questions 4 and 5, but then might not have used the abstract for Report 2 and obtained a data accuracy score of 50% on related Questions 6 and 7. The accuracy rate had to be tied directly to whether or not the abstract was used and thus tied to each reporting/support instance. Variable settings used for this data are is shown in F*igure 3.08*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	8	0	Analysis Accuracy (% Correct)	None	None	8	Left	Scale	Target
AbstractUse	Numeric	8	0	Abstract Use (0 Not Used, 1 Used)	None	None	8	Left	Nominal	Input

Figure 3.08: *Abstract Use and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and abstract use as the grouping variable. This resulted in the statistics shown in *Appendix G*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between abstract use and data analysis accuracy was significance.

Interpretation guide use and data analysis accuracy**.** The researcher needed to determine the relationship between whether or not an interpretation guide was used by the respondent, as indicated by the respondent, and the resultant data analysis accuracy. In order to isolate only instances where an interpretation guide was used or not used, which included data from control group participants and participants who were given a reporting environment with interpretation guides, contents were *only* used from rows where the value of Column Q equaled 1, 6, or 7 (see *Appendix D* for column code book descriptions). The applicable respondent data row contents of Columns IQ and IR were

added to SPSS, followed (underneath) by that of Columns IS and IT, then IU and IV, and then IW and IX. This created two columns and 364 rows of applicable respondent data in order to include data for all four data analysis questions that were answered by each respondent, each with a chance of an interpretation guide being u used or not used. For example, a participant might have used the interpretation guide for Report 1 and obtained a data accuracy score of 100% on related Questions 4 and 5, but then might not have used the interpretation guide for Report 2 and obtained a data accuracy score of 50% on related Questions 6 and 7. The accuracy rate had to be tied directly to whether or not the interpretation guide was used and thus tied to each reporting/support instance. Variable settings used for this data are is shown in F*igure 3.09*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	8	0	Analysis Accuracy (% Correct)	None	None	8	Left	Scale	Target
InterpGuideUse	Numeric	8	0	Interp. Guide Use (0 Not Used, 1 Used)	None	None	8	Left	Nominal	Input

Figure 3.09: *Interpretation Guide Use and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and interpretation guide use as the grouping variable. This resulted in the statistics shown in *Appendix H*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between interpretation guide use and data analysis accuracy was significance.

Support presence and data analysis accuracy. The researcher needed to determine the relationship between whether or not any data analysis support (e.g., footer, abstract, or interpretation guide) was available to the respondent and the resultant data analysis accuracy. The respondent data row contents of Columns IY and IZ were added

to SPSS, followed (underneath) by that of Columns JA and JB, then JC and JD, and then JE and JF (see *Appendix D* for column code book descriptions). This created two columns and 844 rows of respondent data in order to include data for all four data analysis questions that were answered by each respondent, each with a chance of a data analysis support being present or not present. Variable settings used for this data are is shown in F*igure 3.10*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	8	0	Analysis Accuracy (% Correct)	None	None	8	Left	Scale	Target
SupportPresence	Numeric	8	0	Support Presence (0 Not Present, 1 Present)	None	None	8	Left	Nominal	Input

Figure 3.10: *Support Presence and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and support presence as the grouping variable. This resulted in the statistics shown in *Appendix I*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between support presence and data analysis accuracy was significance.

Footer presence and data analysis accuracy. The researcher needed to determine the relationship between whether or not a footer was available to the respondent and the resultant data analysis accuracy. In order to isolate only instances where a footer was used or not used, which included data from control group participants and participants who were given a reporting environment with footers, contents were *only* used from rows where the value of Column Q equaled 1, 2, or 3 (see *Appendix D* for column code book descriptions). The applicable respondent data row contents of Columns IY and IZ were added to SPSS, followed (underneath) by that of Columns JA and JB, then JC and JD,

and then JE and JF. This created two columns and 364 rows of applicable respondent data in order to include data for all four data analysis questions that were answered by each respondent, each with a chance of a footer being present or not present. Variable settings used for this data are is shown in F*igure 3.11*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	8	0	Analysis Accuracy (% Correct)	None	None	8	Left	Scale	Target
FooterPresence	Numeric	8	0	Footer Presence (0 Not Present, 1 Present)	None	None	8	Left	Nominal	Input

Figure 3.11: *Footer Presence and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and footer presence as the grouping variable. This resulted in the statistics shown in *Appendix J*. The t value from the t-test for Equality of Means was used to determine whether the relationship between footer presence and data analysis accuracy was significance.

***Abstract presence and data analysis accuracy*.** The researcher needed to determine the relationship between whether or not an abstract was available to the respondent and the resultant data analysis accuracy. In order to isolate only instances where an abstract was used or not used, which included data from control group participants and participants who were given a reporting environment with abstracts, contents were *only* used from rows where the value of Column Q equaled 1, 4, or 5 (see *Appendix D* for column code book descriptions). The applicable respondent data row contents of Columns IY and IZ were added to SPSS, followed (underneath) by that of Columns JA and JB, then JC and JD, and then JE and JF. This created two columns and 364 rows of applicable respondent data in order to include data for all four data analysis

questions that were answered by each respondent, each with a chance of an abstract being present or not present. Variable settings used for this data are is shown in F*igure 3.12*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	8	0	Analysis Accuracy (% Correct)	None	None	8	Left	Scale	Target
AbstractPresence	Numeric	8	0	Abstract Presence (0 Not Present, 1 Present)	None	None	8	Left	Nominal	Input

Figure 3.12: *Abstract Presence and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and abstract presence as the grouping variable. This resulted in the statistics shown in *Appendix K*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between abstract presence and data analysis accuracy was significance.

Interpretation guide presence and data analysis accuracy. The researcher needed to determine the relationship between whether or not an interpretation guide was available to the respondent and the resultant data analysis accuracy. In order to isolate only instances where an interpretation guide was used or not used, which included data from control group participants and participants who were given a reporting environment with interpretation guides, contents were *only* used from rows where the value of Column Q equaled 1, 6, or 7 (see *Appendix D* for column code book descriptions). The applicable respondent data row contents of Columns IY and IZ were added to SPSS, followed (underneath) by that of Columns JA and JB, then JC and JD, and then JE and JF. This created two columns and 364 rows of applicable respondent data in order to include data for all four data analysis questions that were answered by each respondent, each with a

chance of an interpretation guide being present or not present. Variable settings used for this data are is shown in F*igure 3.13*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	8	0	Analysis Accuracy (% Correct)	None	None	8	Left	Scale	Target
InterpGuidePresence	Numeric	8	0	Interp. Guide Presence (0 Not Present, 1 Present)	None	None	8	Left	Nominal	Input

Figure 3.13: *Interpretation Guide Presence and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and interpretation guide presence as the grouping variable. This resulted in the statistics shown in *Appendix L*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between interpretation guide presence and data analysis accuracy was significance.

Footer format and data analysis accuracy. The researcher needed to determine the relationship between a footer's format, as explored through two formats differing in length and color usage, and respondents' resultant data analysis accuracy. In order to isolate only instances where a footer was present, which included only data from participants who were given a reporting environment with footers, contents were *only* used from rows where the value of Column Q equaled 2 or 3 (see *Appendix D* for column code book descriptions). The applicable respondent data row contents of Columns Q and DK were added to SPSS, and the percentage contents of Column DK were converted into numeric values (e.g., "50 %" was converted to "50") in order to avoid rejected from use as the test variable in SPSS. This created two columns and 60 rows of applicable respondent data, as 60 respondents were given reporting environments featuring footers. Variable settings used for this data are is shown in F*igure 3.14*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	4	0	Analysis Accuracy (% Correct)	None	None	8	Left	Ordinal	Target
FooterFormat	Numeric	8	0	Footer Format (2 Shorter, 3 Longer)	None	None	8	Left	Nominal	Input

Figure 3.14: *Footer Format and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and footer format as the grouping variable. This resulted in the statistics shown in *Appendix M*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between footer presence and data analysis accuracy was significance.

***Abstract format and data analysis accuracy*.** The researcher needed to determine the relationship between a abstract's format, as explored through two formats differing in length and color usage, and respondents' resultant data analysis accuracy. In order to isolate only instances where a abstract was present, which included only data from participants who were given a reporting environment with abstracts, contents were *only* used from rows where the value of Column Q equaled 4 or 5 (see *Appendix D* for column code book descriptions). The applicable respondent data row contents of Columns Q and DK were added to SPSS, and the percentage contents of Column DK were converted into numeric values (e.g., "50 %" was converted to "50") in order to avoid rejected from use as the test variable in SPSS. This created two columns and 60 rows of applicable respondent data, as 60 respondents were given reporting environments featuring abstracts. Variable settings used for this data are is shown in F*igure 3.15*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	4	0	Analysis Accuracy (% Correct)	None	None	8	Left	Ordinal	Target
AbstractFormat	Numeric	8	0	Abstract Format (4 Simpler, 5 Denser)	None	None	8	Left	Nominal	Input

Figure 3.15: *Abstract Format and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and abstract format as the grouping variable. This resulted in the statistics shown in *Appendix N*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between abstract presence and data analysis accuracy was significance.

Interpretation guide format and data analysis accuracy. The researcher needed to determine the relationship between a interpretation guide's format, as explored through two formats differing in length and color usage, and respondents' resultant data analysis accuracy. In order to isolate only instances where a interpretation guide was present, which included only data from participants who were given a reporting environment with interpretation guides, contents were *only* used from rows where the value of Column Q equaled 6 or 7 (see *Appendix D* for column code book descriptions). The applicable respondent data row contents of Columns Q and DK were added to SPSS, and the percentage contents of Column DK were converted into numeric values (e.g., "50 %" was converted to "50") in order to avoid rejected from use as the test variable in SPSS. This created two columns and 60 rows of applicable respondent data, as 60 respondents were given reporting environments featuring interpretation guides. Variable settings used for this data are is shown in F*igure 3.16*.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
Accuracy	Numeric	4	0	Analysis Accuracy (% Correct)	None	None	8	Left	Ordinal	Target
InterpGuideFormat	Numeric	8	0	Interp. Guide Format (6 2-Page, 7 3-Page)	None	None	8	Left	Nominal	Input

Figure 3.16: *Interpretation Guide Format and Data Analysis Accuracy Variable Settings*

The SPSS *Analyze: Compare Means: Independent Samples T-Test* function was then used to conduct an Independent Samples T-Test with a 95% confidence interval, analysis accuracy as the test variable, and interpretation guide format as the grouping variable. This resulted in the statistics shown in *Appendix O*. The *t* value from the t-test for Equality of Means was used to determine whether the relationship between interpretation guide presence and data analysis accuracy was significance.

Crosstabulations with Chi-square. The SPSS *Analyze: Descriptive Statistics: Crosstabs* function was used to conduct Chi-square analyses (*Appendices I-J*). The relationships between independent variables 3-13 (as indicated in *Table 3.07*) and

- (a) respondents' data analysis accuracy and
- (b) respondents' likelihood of using embedded data analysis supports

were examined with Chi-square analyses in order to answer secondary research Questions 5a-6e (see *Table 3.02*). To do this, the 211 respondent data file rows for Columns A, C-E, O-Q, W-Z, DK, IO, JG, and JH (see *Appendix D* for code book definitions) was pasted into SPSS. Variable settings used for this data when examining respondents' data analysis accuracy, (a), are is shown in F*igure 3.17*. Variable settings used for this data when examining respondents' likelihood of using embedded data analysis supports, (b), remained the same as those shown in F*igure 3.17* except the last two variable rows (*Accuracy* and *SupportUse*) had their roles swapped.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
N	Numeric	8	0	#	None	None	8	Left	Scale	None
Veteran	String	16	0	Veteran Status	None	None	16	Left	Ordinal	Input
Role	String	53	0	Role	None	None	53	Left	Ordinal	Input
Perceived	String	19	0	Perceived Data Analysis Proficiency	None	None	19	Left	Ordinal	Input
PD	String	9	0	PD	None	None	9	Left	Ordinal	Input
Courses	String	9	0	Courses	None	None	9	Left	Ordinal	Input
Folder	Numeric	8	0	Folder/Scenario	None	None	8	Left	Nominal	Input
API	Numeric	8	0	API	None	None	8	Left	Scale	Input
EL	String	3	0	English Learner	None	None	3	Left	Ordinal	Input
SocioEc	String	3	0	Socioeconomically Disadvantaged	None	None	3	Left	Ordinal	Input
StudentsWDis	String	3	0	Students with Disabilities	None	None	3	Left	Ordinal	Input
SchoolLevelType	Numeric	8	0	School Level Type (1 Elem., 2 Sec.)	None	None	8	Left	Ordinal	Input
SchoolLevel	Numeric	8	0	School Level (1 Elem., 2 Mid./Jr., 3 High)	None	None	8	Right	Ordinal	Input
Accuracy	String	4	0	Analysis Accuracy (% Correct)	None	None	4	Left	Ordinal	Input
SupportUse	String	4	0	Support Use/Want	None	None	4	Left	Ordinal	Target

Figure 3.17: *Demographics and Data Analysis Accuracy Variable Settings*

For each analyses of a variable's relationship to respondents' data analysis accuracy, (a):

- a single variable (of Variables 3-13) was selected for the crosstab's *Row(s)* value
- the *Data Analysis Accuracy (% Correct)* variable, which was originally derived from Column DK in the data file (see *Appendix D* for code book definition), was selected for the crosstab's *Column(s)* value, and
- *Chi-square* was selected from the *Crosstabs: Statistics* options.

For each analyses of a variable's relationship to respondents' likelihood of using embedded data analysis supports, (b):

- a single variable (of Variables 3-13) was selected for the crosstab's *Row(s)* value
- the *Support Use/Want* variable, which was originally derived from Column IO in the data file (see *Appendix D* for code book definition), was selected for the crosstab's *Column(s)* value, and
- *Chi-square* was selected from the *Crosstabs: Statistics* options.

Since crosstabulations with Chi-square analyses were conducted for 2 relationship types (a and b, as outlined above) for each of the for 11 independent variables described above,

22 crosstabulations with Chi-square analyses were conducted. Results from these analyses are featured in *Appendices I-J*. The Chi-square tests allowed the researcher to determine if any relationships between crosstabulated variables were significant as opposed to random variation. The significance value (Asymp. Sig.) was used to determine the significance value of the relationships, with the lower the value the more likely the two variables were deemed related, and with significance values less than 0.05 deemed significant.

Assumptions

The study was created under the assumption of this fact to be true: educators' data analyses impact students when these analyses are used to inform decisions made specifically to impact students. Thus data analysis *errors* made within the data analysis step of data-informed decision-making have the potential to negatively impact students and therefore constitute a problem that needs to be remedied. However, the assumption that served to inspire this study was not the only assumption made.

Assumptions about the study population included that respondents would make reasonable attempts to answer the four data analysis questions – Questions 4-7 – correctly, but they would not necessarily answer the questions to the best of their abilities. Because most survey completion sessions were conducted at the end of the school day, which meant at the end of each participant's work day, it was reasonable to assume respondents were tired, which is not conducive to data analysis accuracy. For example, fatigue at the end of a workday can cause a significant decline in interpretation accuracy (Krupinski & Berbaum, 2010). However, the times when these survey sessions were conducted – when staff members were not teaching – were also the time these

educators would be most likely to have the time to conduct their real-life analyses of student data. Thus these were the ideal times to conduct survey sessions. Nonetheless, steps were taken to reduce other factors that might artificially reduce analysis performance. For example, test anxiety and worry over potential negative results from a test can cause testers to perform poorly in testing situations (Zeidner, 1998). Thus the fact that responses were completely anonymous and could not be tied back to the individuals taking the survey was a fact that was included on the Informed Consent Form but also verbally stressed to all participants when the study was introduced.

Another assumption about the study population was that respondents would be honest in their responses. For example, it was assumed that a first year teacher would be honest in indicating he or she had been teaching for no more than one year on Question 1 as opposed to making the "20 or more years" selection. Nonetheless, steps were taken to best ensure such honestly. For example, the study was voluntary and its voluntary nature was stressed on the Informed Consent Form each participant signed but also verbally to all participants when the study was introduced. Respondents were told there would be no negative repercussions if they opted not to participate, and they were told they could withdraw at any time, even after beginning the survey. In this way any educators not interested in making the honest efforts needed to participate could easily abstain. In addition, the researcher expressed deep gratitude for participants' time and feedback at the start of the survey and stressed the impact the study results are likely to have on educators and students in the future. This atmosphere of gratitude likely helped participants to know they were appreciated, which likely increased rather than decreased their chances of making honest efforts to complete the survey with legitimacy. Likewise,

educators are in a profession focused on helping students, so the fact that results would impact educators and students was also likely to increase the fidelity of their answers. The also researcher circulated the room in the same manner teachers do in their classrooms during test sessions, which helped to maintain silence and seriousness, while also avoiding diffusion of treatment.

An assumption about the study design was that that 211 sample size would render educators demonstrative of all participant characteristics featured in *Table 3.04*. In an effort to have all of these participant characteristics manifested in the study sample, the researcher conducted a priori two-tailed t-test calculating the difference between two independent means to determine ideal sample size. The priori two-tailed t-test resulted in a recommended sample size of at least 210 educators. However, the researcher also conducted an F-test linear multiple regression analysis, fixed model, R^2 deviation from zero. This priori F-test resulted in a recommended sample size of at least 153 educators. However, since the 210 sample size resulting from the two-tailed t-test was greater than 153, responses from 211 participants were collected for the study in order to exceed even the more rigorous recommendation. See *Chapter 3: Research Method: Research Method and Design* for details on the regression analyses that was also applied. Fortunately, the 211 sample size successfully rendered educators demonstrative of all participant characteristics featured in *Table 3.04*. This allowed the sample size to be generalized to the education population at large.

Limitations

The study dealt exclusively with educators and their use of data system reports and resources in an isolated setting. Thus, to maintain external validity, study findings

may not be applied to inferences concerning non-educators, such as parents, students, or politicians. Likewise, in consideration of the potential impact of interaction of setting and treatment, no generalizations of data analyses may be made of analysis environments that are not report-based, such as data analyses made based on data group discussions or based on an explanation heard by a data coach.

Overcoming threats to external validity. Both external validity and construct validity involve making generalizations, but a distinction exists in the types of generalizations made. Regarding construct validity, the study's measurements clearly assessed data analysis accuracy – the study's "label" – and were not more appropriate for another topic. Regarding external validity, the study was accurately applied to inferences concerning other educators who interact with similar yet different data and data reports. This was the case because external validity threats are risked when sample data is used to draw conclusions concerning other people, settings, or time periods (Vogt, 2006).

The threats to external validity are interaction of selection and treatment, interaction of setting and treatment, and interaction of history and treatment (Black, 1999; Gall, Borg, & Gall, 2007). To avoid the first of these, many educators of varied roles were included in the study, and generalizations about educators outside of those included in the study were not made. For example, inferences may only be made to veteran teachers if they were also thoroughly represented in the study; likewise, inferences may not be made to parents or students using data system reports. To circumvent the second of these threats, no generalizations of data analyses were made of analysis environments that are not report-based, such as data analyses made based on data group discussions or based on an explanation heard by a data coach. To avoid the last of these threats, no

generalizations were made about past or future data analyses that could be altered by history's impact on the treatment. For example, if an educator uses data reports with the analysis guidance of footers, abstracts, and interpretation guides for one year and then suddenly uses the same reports without these supports, it is *not* supposed his or her data analyses will be as poor as those who used support-free reports in the study, as his or her understanding of the data's proper analyses will likely have been impacted by the regular use of support-embedded reports.

Overcoming threats to internal validity. The study incorporated cause and effect inferences, as its researcher sought to assess the impact analysis supports in data systems have on the accuracy of educators' data analysis while using the reports. Thus internal validity and its threats were aspects that were considered. The threats to internal validity are lengthy: history, maturation, regression, selection, mortality, diffusion of treatment, compensatory/resentful demoralization, compensatory rivalry, testing, and instrumentation (Black, 1999). Some of these threats were not relevant to this study: maturation, regression, mortality, compensatory/resentful demoralization, compensatory rivalry, testing, and instrumentation. For example, maturation was not an issue, as no more than 20 minutes passed between the start and end of each study session, and thus participants did not significantly age during survey completion. As another example, compensatory/resentful demoralization and compensatory rivalry were not concerns because participants were not aware of the treatment other participants were receiving during the study and/or how they differed.

To avoid the internal validity threats that could relate to the study, all groups were exposed to the same external events and participants were selected randomly while still

being selecting from varied sites and school levels. These steps helped to circumvent the threats of history and selection. Diffusion of treatment, however, was a concern. The reports participants worked with during the study varied slightly from those used by others in the room in terms of the supports that were on the reports or accompanied the reports, though participants were asked to work independently and not interact, and thus did not see the differences between report environments or have their responses affected. However, most teachers work in isolation from their colleagues for the majority of the workday and thus could have been eager to interact when outside the classroom. To avoid diffusion of treatment, the researcher circulated the room in the same manner teachers do in their classrooms during test sessions, and the need for silence was addressed as the study was introduced and maintained during the study.

Delimitations

Although the study's scope concerned guidance that computer-based data systems can provide within reports they are used to generate and within the data systems themselves, participants used reports and supports that can come from a data system as opposed to actually using a data system on a computer. Viewing a data system's report on the computer versus printed can negatively impact how it is interpreted; for example, someone who correctly interprets a printed report can make mistakes when scrolling is involved (Hattie, 2010; Leeson, 2006). Also, technology can prevent someone from demonstrating a skill when he or she lacks computer familiarity (Bennett & Gitomer, 2009; Horkay, Bennett, Allen, Kaplan, & Yan, 2006). For example, technology problems such as outdated hardware, inadequate bandwidth, system freezes, and use of computers outside of the teaching profession influence teachers' success using a data system

(Rennie Center for Education Research and Policy, 2006; Underwood et al., 2008). In order to prevent such variables as technical skills and Internet conditions impacting study results, the study involved printed data system reports and handouts. This way the researcher and did not have to wonder – as those reading the study findings will not have to wonder – if analysis struggles were due to the study's report and support environments or merely due to technical struggles or Internet problems.

Another delimitation of the study concerned data-informed decision-making and behavioral economics. This study related to improving the accuracy of educators' data analyses, which is enacted in the thought portion – or "data-informed" portion – of data-informed decision-making. According to the ideomotor effect, primed thoughts then prime one's decisions or behavior (Kahneman, 2011). Thus the data-informed thoughts are believed to influence decision-making. Nonetheless, this study did not explore the decision-making that results from the data-informed thoughts.

Many behavioral economics dimensions can be manipulated to improve data-informed decision-making. For example, the process of thinking and deciding is influenced by behavioral economics facets such as priming, biases, heuristics, prototypes, judgments, anchoring, and framing (Kahneman, 2011). Even seemingly insignificant differences in how content is arranged can mean a significant difference in the decisions people make based on that content (Thaler & Sunstein, 2008). However, this study was concerned only with the reporting environments generated by the online data system, as the study's purpose lay in finding ways data systems can be improved to facilitate improved data analyses. Thus conditions outside of those that can be controlled within a data system were not manipulated or used as variables in the study.

The study's final delimitation related to the analysis support formats that were investigated. There is an abundance of research concerning general recommendations and best practices for report footers and for supplemental documentation such as abstracts and interpretation guides. For example, too much information or text can overwhelm users and cause them to miss higher level implications, so links to supplemental information can allow users to "drill down" to information that is not the focus of the report (VanWinkle, Vezzu, & Zapata-Rivera, 2011). The three analysis supports used as independent variables in this study thus conformed to the wealth of existing research concerning their format. However, there were finer aspects of format that had not been investigated in terms of specific impact on educators' data analyses. Thus this study's exploration of format, through the two differing formats that were used for each analysis support, was not an investigation of whether or not "format matters" in regards to these tools. Rather, since it is already accepted the format of such tools *does* matter, generally-similar yet slightly-dissimilar formats were investigated, namely concerning length and color usage, to explore finer points of analysis support format. For example, it is already known longer paragraphs discourage users from reading them, with some research indicating passages with short paragraphs receive twice as much attention as those with longer paragraphs (Outing & Ruel, 2006). Thus this study's reports bearing footers did not feature a two-line footer in one reporting environment and then a half-page footer in the second reporting environment; such differences would be extreme and the likely outcome already known. Rather, the difference between footer length in this study was more subtle: 39 words versus 58 words for Report 1, and 34 words versus 42 words for Report 2 (see *Chapter 3: Research Method: Materials/Instruments: Handouts* for specific

details on handout differences, and see *Appendix C* for the actual handouts). Thus findings could be used to determine more specific recommendations for analysis supports if those findings were found to be significant. In the case of this study, unlike the other support-related variables the study investigated, the format-related differences found were not deemed significant.

Ethical Assurances

Deliberate measures were taken to ensure the study adhered to ethical practices, such as protection from harm, informed consent, right to privacy, and honesty with professional colleagues. The researcher took key steps at each stage of the doctoral process to apply the care and integrity needed to meet the ethical standards of scientific research. For example:

Plagiarism. All work submitted in relation to the study was the author's own or else properly cited. This means every portion of the dissertation includes proper citations throughout, conforming to guidelines found in the 6th edition of *Publication Manual of the American Psychological Association* (APA, 2001). This includes the *Self-plagiarism* section of Chapter 1 in the publication.

Risk assessment. Per Title 45 CFR 46.102(i) of Federal Regulations, the study needed to involve minimal risk to those involved in the study, meaning that it was not greater than risks normally encountered in everyday life or during routine examinations (Office of Human Subjects Research National Institutes of Health, 2005). To be sure the study conformed to this specification, the researcher ensured the testing environment was safe. For example, any cables or cords running through the computer lab were safely kept out of walking areas so no one tripped, and rooms were arranged for easy passage to and

from seats. Psychological factors were also considered, as response data was kept anonymous so no educator or institute could feel embarrassment or incur repercussions as a result of involvement in the study.

Informed consent. The researcher exercised a proactive approach to informed consent. For example, consider the NCU Informed Consent Checklist item, "Statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time without penalty or loss of benefits, to which the individual is otherwise entitled" (Northcentral University, 2011, p. 1). While this information was featured on the form, it was ethically responsible to also frontload those helping to recruit participants. Since the researcher was organizing participation sessions at schools through their administrators, she took steps to make sure there was no miscommunication between each principal and his or her staff that might leave participants to guess their participation was required. For example, the researcher offered sample verbiage for principals' emails and fliers to staff so busy principals could accurately communicate the voluntary nature of participation. Because of the potential for misunderstanding, the researcher was also extra careful to clearly communicate participants' options for backing out without negative consequences by delivering this message to participants at the onset of the session in written and verbal format.

Privacy, confidentiality, and data handling. The researcher adhered to all APA Ethics Code standards, specifically Standards 8.01-8.09 dealing with the treatment of humans and animals (APA, 2001). The researcher also selected and used tools to facilitate the protection of confidentiality. For example, she used the Google Docs Form

feature for a survey to collect participant responses without human interaction. This tool automatically assigns an anonymous, unique identifier (ID) to each record/row of response data. These IDs were thus used in a complete absence of participant names or employee numbers. Results were thus kept anonymous, as there was no record that tied responses back to specific participants.

Study design and reporting. The researcher also took steps to protect the integrity of results reporting and its ability to be applied to real world practice. For example, the Google Docs Form "required question" setting was assigned to each survey question to eliminate the risk of response bias resulting from nonresponses on the survey. Results were tabled, graphed to check for normal distribution, and tested to see if they were considered statistically significant. A descriptive analysis containing the means, standard deviations, and score ranges was then prepared in relation to the independent and dependent variables. See *Chapter 4: Results* for details such as the varying significance levels (p) used for different types of research questions. Straightforward categorical scales in the form of correct/incorrect were used for analysis questions, as the answers were clearly right or wrong.

Overcoming threats to construct validity. Data systems can contain footers on the reports they are used to generate, as well as abstracts and interpretation guides via links that accompany these reports in the data system and can also be printed to accompany printed reports. When applied to this study, construct validity described the degree to which inferences made based on the chosen measurement instrument may be applied to the theory concerning improving data analysis in real environments through data system modifications, such as the inclusion of footers, abstracts, or interpretation

guides with analysis guidance directly related to data contained in a data system report. For this study to have construct validity, the questions participants answered had to be appropriate measures of data analysis competency, and performance in answering them had to clearly reflect the impact of the variables used in the study. For example, the impact of independent variables such as particular forms of analysis guidance on the dependent variable were appropriately measured.

Threats to construct validity include poor preoperational explanation of constructs, mono-operational bias, mono-method bias, interaction of different treatments and/or testing and treatment, not factoring in untended consequences on constructs, confounding constructs, and social threats (Vogt, 2006). There are various measures this study incorporated to avoid these. For example, the study:

- had constructs that were clearly defined,
- captured the full scope of the program by conducting the same experiment at a variety of school sites and with varied educators,
- used multiple questions in the measurement tool,
- accounted for how treatments interacted with one another as well as with the measurement itself,
- appropriately considered unintended consequences,
- and labeled experiment elements properly.

To sidestep natural social tendencies, steps were taken to reduce the impact of hypothesis guessing, evaluation apprehension, and experimenter expectancies.

Overcoming threats to external validity. Both external validity and construct validity involve making generalizations, but a distinction exists in the types of

generalizations made. Regarding construct validity, the study's measurements clearly assessed data analysis accuracy – the study's "label" – and were not more appropriate for another topic. Regarding external validity, the study was accurately applied to inferences concerning other educators who interact with similar yet different data and data reports. This was the case because external validity threats are risked when sample data is used to draw conclusions concerning other people, settings, or time periods (Vogt, 2006).

The threats to external validity are interaction of selection and treatment, interaction of setting and treatment, and interaction of history and treatment (Black, 1999; Gall, Borg, & Gall, 2007). To avoid the first of these, many educators of varied roles were included in the study, and generalizations about educators outside of those included in the study were not made. For example, inferences may only be made to veteran teachers if they were also thoroughly represented in the study; likewise, inferences may not be made to parents or students using data system reports. To circumvent the second of these threats, no generalizations of data analyses were made of analysis environments that are not report-based, such as data analyses made based on data group discussions or based on an explanation heard by a data coach. To avoid the last of these threats, no generalizations were made about past or future data analyses that could be altered by history's impact on the treatment. For example, if an educator uses data reports with the analysis guidance of footers, abstracts, and interpretation guides for one year and then suddenly uses the same reports without these supports, it is *not* supposed his or her data analyses will be as poor as those who used support-free reports in the study, as his or her understanding of the data's proper analyses will likely have been impacted by the regular use of support-embedded reports.

Overcoming threats to internal validity. The study incorporated cause and effect inferences, as its researcher sought to assess the impact analysis supports in data systems have on the accuracy of educators' data analysis while using the reports. Thus internal validity and its threats were aspects that were considered. The threats to internal validity are lengthy: history, maturation, regression, selection, mortality, diffusion of treatment, compensatory/resentful demoralization, compensatory rivalry, testing, and instrumentation (Black, 1999). Some of these threats were not relevant to this study: maturation, regression, mortality, compensatory/resentful demoralization, compensatory rivalry, testing, and instrumentation. For example, maturation was not an issue, as no more than 20 minutes passed between the start and end of each study session, and thus participants did not significantly age during survey completion. As another example, compensatory/resentful demoralization and compensatory rivalry were not concerns because participants were not aware of the treatment other participants were receiving during the study and/or how they differed.

To avoid the internal validity threats that could relate to the study, all groups were exposed to the same external events and participants were selected randomly while still being selecting from varied sites and school levels. These steps helped to circumvent the threats of history and selection. Diffusion of treatment, however, was a concern. The reports participants worked with during the study varied slightly from those used by others in the room in terms of the supports that were on the reports or accompanied the reports, though participants were asked to work independently and not interact, and thus did not see the differences between report environments or have their responses affected. However, most teachers work in isolation from their colleagues for the majority of the

workday and thus could have been eager to interact when outside the classroom. To avoid diffusion of treatment, the researcher circulated the room in the same manner teachers do in their classrooms during test sessions, and the need for silence was addressed as the study was introduced and maintained during the study.

Mistakes and negligence. To avoid mistakes and negligence, the researcher regularly referred to procedural texts and Northcentral resources such as the handbooks and Dissertation Center. In the event of any mistakes or negligence, the researcher would have immediately sought the counsel of her mentor, responded accordingly, referred to Shapiro and Smith (2011) for added input. In her dissertation the researcher would also have been honest and open about any mistakes so readers and future research may be thoroughly informed. However, such steps were not necessary as no mistakes or negligence occurred.

IRB approval. The researcher completed the Collaborative Institutional Training Initiative (CITI) course, studied related literature such as Fiore (2011), and reviewed the Dissertation Center’s Institutional Review Board (IRB) Information section, which includesd the IRB Application. The researcher did not have any large ethical concerns for the intended research topic but always proceeded with caution in all areas nonetheless. Northcentral University IRB approval was obtained prior to the collection of any data for this study.

Summary

While a doctor isn’t present to explain an over-the-counter medication’s use, medicine bought in a store comes with a detailed label outlining its purpose, ingredients, dosage instructions, and dangers. It would be negligent to sell medicine without such

guidance on how to use it wisely, as this would risk the lives of those the medicine is used to treat (Brown-Brumfield & DeLeon, 2010, DeWalt, 2010). Meanwhile, educators are using data to treat students, yet they are operating without the data-equivalent to over-the-counter medicine: reports generated in data systems typically contain insufficient supports such as labeling or supplemental documentation to guide users in the data's use. The vast majority of stakeholders who use student data are not trained statisticians, and they need the data they view to be accompanied with additional information to teach them how to understand and use the data (DQC, 2009). Yet educators are using data systems and data system reports that do not feature data analysis guidance to help educators use the data appropriately – much like ingesting medicine from an unmarked or marginally marked container. Hampton (2007), Qin et al. (2011), and Clay (2012) offered or called for label recommendations similar to those recommended by the FDA for over-the-counter medication labels. Label conventions can result in improved understanding on non-medication products, as well, if they are included (Hampton, 2007; Qin et al., 2011).

Research on aspects of report format and system support that can improve analysis accuracy is scarce (Goodman & Hambleton, 2004). Research that was devoted to data system and report format focuses on participants' preferences and participants' perceived value of supports as opposed to measuring supports' actual impact on interpretation. This study was used to examine exactly how effective varied analysis supports, appropriate for inclusion in the data systems being used for data analyses, are in improving data analysis accuracy. The findings of this study contributed to literature in the field by helping to identify how data systems can best help increase data analysis

accuracy by providing analysis support within data systems and their reports. This means the findings have the potential to benefit students, who deserved to have this potential source of help explored.

Chapter 4: Findings

The *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study investigated the problem of educators making data analysis errors impacting students while data systems and reports do not include analysis help, whereas it was undecided whether adding supports to data systems can reduce the number of analysis errors. Data-informed decisions can improve learning (Sabbah, 2011; Underwood, Zapata-Rivera, & VanWinkle, 2010; Wohlstetter, Datnow, & Park, 2008). Educators worldwide test students, distribute score reports, and expect stakeholders to make improvements based on these reports (Hattie & Brown, 2008). Most educators have access to data systems to generate and analyze score reports (Aarons, 2009; Herbert, 2011).

Unfortunately, educators do not use this data correctly, and there is clear evidence many users of data system reports have trouble understanding the data (Hattie, 2010; National Research Council, 2001; Wayman et al., 2010; Zwick et al., 2008). For example, in a national study of districts known for *strong* data use, teachers incorrectly interpreted 52% of data (U.S. Department of Education Office of Planning, Evaluation and Policy Development [USDEOPEPD], 2009). Few teacher preparation programs cover topics like assessment data literacy (Halpin & Cauthen, 2011; Stiggins, 2002), most people analyzing data received *no* training to do so (DQC, 2009; Few, 2008), and human biases compromise judgment and complicate decision-making processes (Kahneman, 2011).

Data use impacts students, and misunderstandings when using data systems can cripple data use in school districts (Wayman, Cho, & Shaw, 2009). Yet labeling and tools within data systems to assist analysis are uncommon, even though most educators

analyze data alone (USDEOPEPD, 2009). There is a clear need for research identifying how reports can better facilitate correct interpretations by its users (Goodman & Hambleton, 2004; Hattie, 2010). The power of data systems that generate these reports will not be realized until researchers contribute to improving data system design to improve analysis (DQC, 2011).

The *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study was used to determine the degree to which three forms of data system-embedded data analysis support can improve the accuracy of educators' data analyses. This chapter contains the study's findings, organized around the study's primary research questions and hypotheses. First the results are reported with descriptive information but otherwise without discussion. Next an evaluation of findings includes interpretation of the results and speculation of their implications. The chapter's key findings are then summarized.

Results

Tables 4.01-4.14 contain results calculated within the data file in the manner described at length in *Chapter 3: Research Method*. When these tables and this section refer to:

- *supports*, they are referring to (a) any support, combining the supports that follow as b-d; (b) footer; (c) abstract; or (d) interpretation guide.
- *support use*, they are referring to instances in which respondents indicated they (a) used the available support or (b) would have used a support, as was a response option for control group participants who did not receive any supports. Note the support use refers to a percent of *instances* and not a percent of *participants*. For

Table 4.01: *All Report Environments*

	Participants		Support Use	Data Analysis Accuracy (% Correct)		
Report Environment	*n*	%	% Used/Wanted	Did Not Use Any Support	Regardless of Support Use	Used Available Support
Any Report Environment (All 211 Respondents)	211	100%	62%	11%	26%	39%

Table 4.02: *Each Report Environment*

	Participants		Support Use	Data Analysis Accuracy (% Correct)				
Report Environment	*n*	%	% Used/Wanted	Would Not Have Used Support	Would Have Used Support	Did Not Use Available Support	Regardless of Support Use	Used Available Support
Plain Report (Control Group)	**31**	**15%**	**87%**	**13%**	**11%**	**n/a**	**11%**	**n/a**
Report with Shorter Footer	30	14%	75%	n/a	n/a	27%	36%	33%
Report with Longer Footer	30	14%	70%	n/a	n/a	6%	32%	40%
Report with Any Footer	**60**	**28%**	**73%**	**n/a**	**n/a**	**15%**	**34%**	**37%**
Plain Report + Less Dense Abstract	30	14%	53%	n/a	n/a	11%	21%	31%
Plain Report + Denser Abstract	30	14%	47%	n/a	n/a	9%	24%	36%
Report with Any Abstract	**60**	**28%**	**50%**	**n/a**	**n/a**	**10%**	**23%**	**33%**
Plain Report + 2-Page Interpretation Guide	30	14%	52%	n/a	n/a	0%	32%	48%
Plain Report + 3-Page Interpretation Guide	30	14%	52%	n/a	n/a	3%	28%	48%
Report with Any Interpretation Guide	**60**	**28%**	**52%**	**n/a**	**n/a**	**2%**	**30%**	**48%**
Report with Any Support	**180**	**85%**	**58%**	**n/a**	**n/a**	**8%**	**29%**	**39%**

Table 4.03: *Survey Questions Involving Data Analysis*

	Support Use	Data Analysis Accuracy
Question / Report (Mean 26%)	% Used/Wanted	% Correct
Question 4 (Report 1)	n/a	29%
Question 5 (Report 1)	n/a	28%
Report 1 (Questions 4 & 5)	**72%**	**28%**
Question 6 (Report 2)	n/a	21%
Question 7 (Report 2)	n/a	27%
Report 2 (Questions 6 & 7)	**53%**	**24%**

Table 4.04: *School Level Type*

	Participants		Support Use	Data Analysis Accuracy
School Level Type (2 Total)	*n*	%	% Used/Wanted	% Correct
Elementary	132	63%	64%	26%
Secondary	79	37%	59%	27%

Table 4.05: *School Level*

	Participants		Support Use	Data Analysis Accuracy
School Level (3 Total)	*n*	%	% Used/Wanted	% Correct
Elementary	132	63%	64%	26%
Middle/Junior High	47	22%	48%	25%
High School	32	15%	75%	30%

Table 4.06: *Academic Performance*

2012 Growth Academic Performance Index (API)	Participants		Support Use	Data Analysis Accuracy
(828 Mean)	*n*	%	% Used/Wanted	% Correct
677	32	15%	75%	30%
794	33	16%	47%	18%
815	24	11%	65%	25%
827	14	7%	50%	41%
847	22	10%	68%	24%
891	28	13%	57%	28%
893	16	8%	75%	28%
895	31	15%	71%	31%
916	11	5%	41%	7%

Table 4.07: *English Learner Population*

% of Site's Students Who Are English Learners	Participants		Support Use	Data Analysis Accuracy
(29% Mean)	*n*	%	% Used/Wanted	% Correct
8%	16	8%	75%	28%
10%	31	15%	71%	31%
16%	11	5%	41%	7%
27%	22	10%	68%	24%
30%	33	16%	47%	18%
33%	24	11%	65%	25%
38%	32	15%	75%	30%
45%	14	7%	50%	41%
46%	28	13%	57%	28%

Table 4.08: *Socioeconomically Disadvantaged Population*

% of Site's Students Who Are Socioecon. Disadvantaged	Participants		Support Use	Data Analysis Accuracy
(52% Mean)	*n*	%	% Used/Wanted	% Correct
22%	11	5%	41%	7%
23%	31	15%	71%	31%
31%	16	8%	75%	28%
43%	28	13%	57%	28%
56%	22	10%	68%	24%
61%	57	27%	54%	21%
78%	46	22%	67%	33%

Table 4.09: *Students with Disabilities Population*

% of Site's Students with Disabilities	Participants		Support Use	Data Analysis Accuracy
(10% Mean)	*n*	%	% Used/Wanted	% Correct
5%	16	8%	75%	28%
8%	28	13%	57%	28%
9%	38	18%	59%	31%
10%	33	16%	59%	18%
11%	33	16%	47%	18%
12%	32	15%	75%	30%
13%	31	15%	71%	31%

Table 4.10: *Veteran Status*

Length of Time Working as an Educator (e.g., Teacher or Administrator) for Students under 19 Years of Age	Participants		Support Use	Data Analysis Accuracy
	n	%	% Used/Wanted	% Correct
Less than 1 Year	2	1%	75%	25%
Minimum of 5 Years	20	9%	70%	35%
Minimum of 10 Years	33	16%	67%	32%
Minimum of 15 Years	67	32%	63%	28%
Minimum of 20 Years	89	42%	58%	21%

Table 4.11:: *Role*

	Participants		Support Use	Data Analysis Accuracy
Best Description of Current Position	*n*	%	% Used/Wanted	% Correct
Teacher	199	94%	63%	26%
Colleague Coach (e.g., Teacher on Special Assignment)	2	1%	25%	25%
Site/School Administrator	8	4%	56%	19%
District Administrator	2	1%	100%	75%

Table 4.12: *Perceived Data Analysis Accuracy*

Perceived Proficiency at Analyzing Student Performance Data	Participants *n*	Participants %	Support Use % Used/Wanted	Data Analysis Accuracy % Correct
Very Proficient	45	21%	72%	27%
Somewhat Proficient	139	66%	61%	27%
Not Proficient	22	10%	57%	23%
Far from Proficient	5	2%	30%	10%

Table 4.13: *Professional Development (PD)*

PD Obtained within Past Year, Specifically Focused on Learning How to Correctly Interpret Student Data	Participants		Support Use	Data Analysis Accuracy
	n	%	% Used/Wanted	% Correct
0 Hours	87	41%	58%	23%
Minimum of 1 Hour	48	23%	63%	26%
Minimum of 2 Hours	39	18%	72%	30%
Minimum of 5 Hours	19	9%	71%	22%
Minimum of 8 Hours	18	9%	53%	36%

Table 4.14: *Graduate Educational Measurement Courses*

Graduate-Level Courses Taken, Specifically Dedicated to Educational Measurement	Participants		Support Use	Data Analysis Accuracy
	n	%	% Used/Wanted	% Correct
0 Courses	100	47%	55%	23%
Minimum of 1 Course	51	24%	70%	30%
Minimum of 2 Courses	35	17%	73%	29%
Minimum of 3 Courses	11	5%	64%	25%
Minimum of 4 Courses	14	7%	61%	27%

example, in *Table 4.01*, 62% of study participants indicated they used supports 62% of the time. This is different than saying 62% of *participants* used or wanted supports, as a single respondent might have used supports only 50% of the time, such as using the footer on Report 1 but not the footer on Report 2.

- *data analysis accuracy*, they are referring to the mean value of participants' percent correct scores earned when answering Questions 4-7 measuring data analysis accuracy.

The results featured in *Tables 4.01-4.14* are organized around the study's research questions, which follow. Research Questions were comprised of Q1-Q3b, which constituted the study's seven primary research questions, and Q4a-Q6e, which constituted the study's 11 secondary research questions serving the sole role of informing implications addressed by the primary research questions.

Q1. Research Question Q1 was asked as follows:

- What impact does data analysis guidance accompanying a data system report in the form of footer, abstract, or interpretation guide have on how frequently educators draw accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that accompanying a report with a support containing analysis guidance in the form of footer, abstract, or interpretation guide would not have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.
- The alternative hypothesis was that accompanying a report with a support containing analysis guidance in the form of footer, abstract, or interpretation

guide would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

The null hypothesis ($H1_0$) was rejected and the alternative hypothesis was accepted ($H1_a$) for Q1 based on the study results reported below. Accompanying a report with a support containing analysis guidance in the form of footer, abstract, or interpretation guide had a significant, positive impact on the frequency of accurate conclusions educators drew concerning student achievement data. This finding is explained in the remainder of this Q1 section.

Table 4.01 features results for all 211 study participants, who indicated they used supports 62% of the time. When respondents did not use any supports, their data analysis accuracy was 11%. All 211 participants, regardless of support use, averaged a data analysis accuracy of 26%. In cases where respondents indicated they used an available support, data analysis accuracy was 39%. In terms of relative and absolute differences, educators' data analyses were 264% more accurate (with an 18 percentage point difference) when any one of the three supports was present and 355% more accurate (with a 28 percentage point difference) when respondents specifically indicated having used the support (see *Figure 4.01*). See *Figure 4.02* for a visual representation of supports' impact on educators' data analyses.

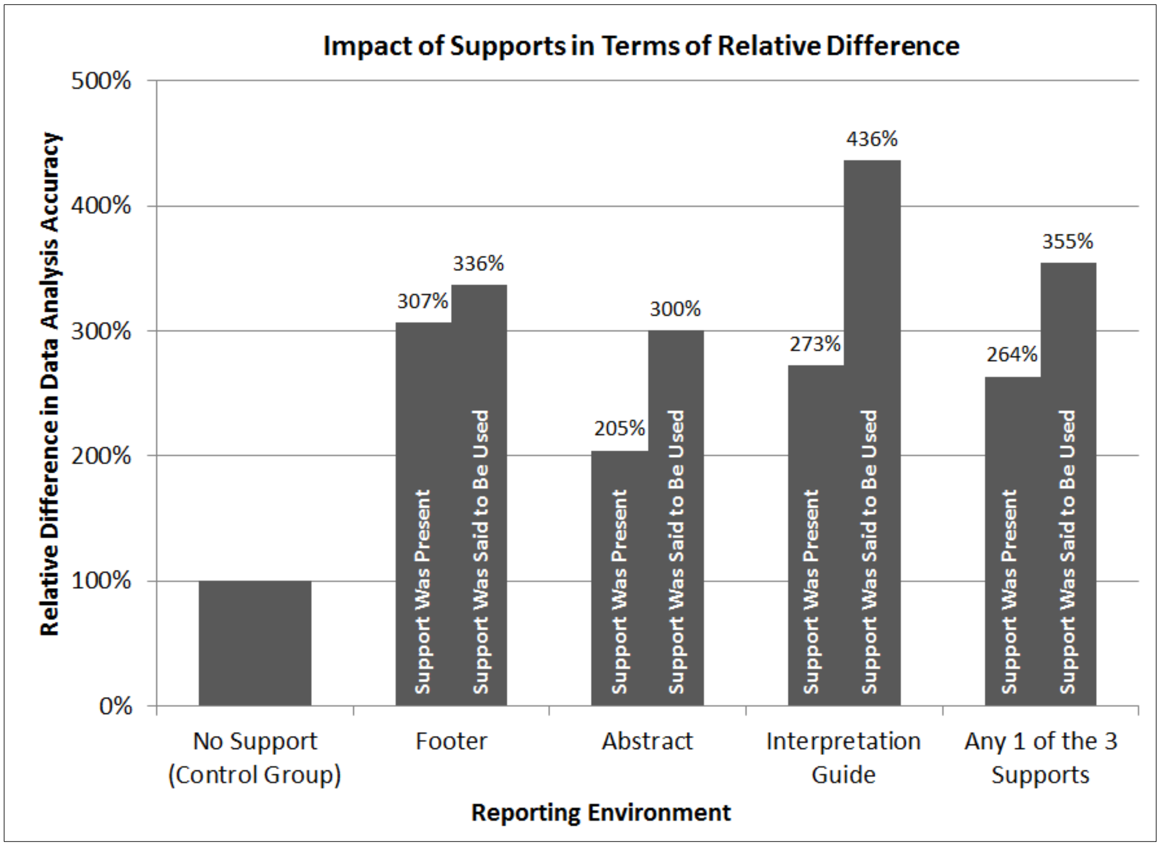


Figure 18: *Impact of Supports in Terms of Relative Difference*

Table 4.02 features results shows for the 31 control group participants, who did not receive any supports, who constituted 15% of the total 211-participant sample. 87% of these participants who had no access to supports indicated they would have used the added support if they had it. Of the 31 control group participants who indicated they would *not* have used the added support, data analysis accuracy was 13%. Of the 31 control group participants who indicated they *would* have used the added support, data analysis accuracy was 11%. All 31 control group participants, regardless of whether or not they wanted supports, averaged a data analysis accuracy of 11%.

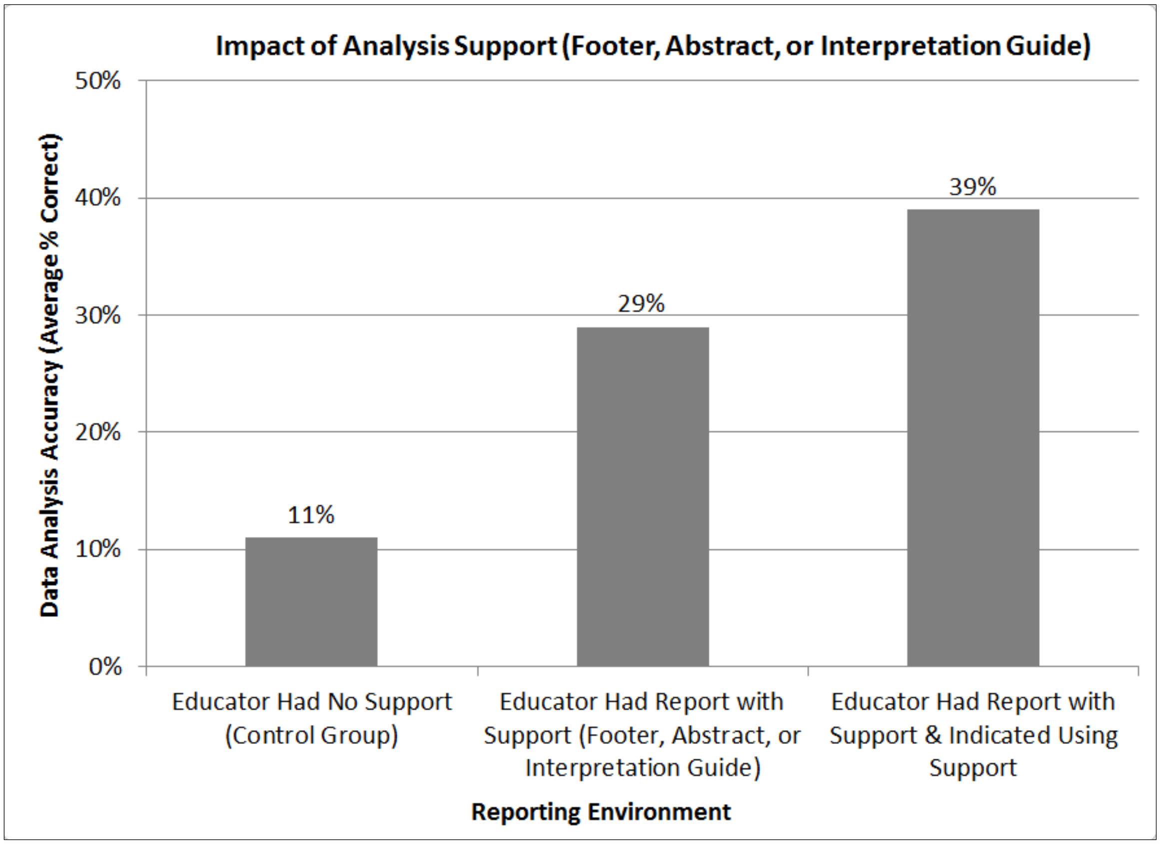


Figure 19: *Impact of Analysis Support (Footer, Abstract, or Interpretation Guide)*

Table 4.02 also features results shows for the 180 participants who received reporting environments containing supports: 60 received footers, 60 received abstracts, and 60 received interpretation guides. These 180 participants constituted 85% of the total 211-participant sample. These participants who had access to report supports indicated they used the supports 58% of the time. When these respondents had supports yet indicated they did not use the supports, their data analysis accuracy was 8%. All 180 participants with supports, regardless of support use, averaged a data analysis accuracy of 29%. In cases where respondents indicated they used the available support, data analysis accuracy was 39%.

Table 4.03 features results by data analysis question on the survey in order to address any questions about whether there was an imbalance in the questions used to measure the data analysis accuracy noted in this section and others. *Table 4.03* features results for all 211 study participants, who indicated they used Report 1 supports 72% of the time, which contributed to their answers of Questions 4 and 5, and used Report 2 supports 53% of the time, which contributed to their answers of Questions 6 and 7. Report 1 was graphical and related to an assessment considered higher stakes than the Report 2 assessment, which was reported in tabular format. Participants' data analysis accuracy was 29% on Question 4 and 28% on Question 5, with an average data analysis accuracy of 28% for Report 1 questions. Participants' data analysis accuracy was 21% on Question 6 and 27% on Question 7, with an average data analysis accuracy of 24% for Report2 questions.

An Independent Samples T-Test (see *Appendix E*) was used to determine whether the supports' impact on educators' data analysis accuracy was significant. This test first compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who used the support and those who did not). As indicated in *Appendix E*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.000. This value was less than 0.10, suggesting the variable groups had unequal variances. Consistent with Levene's Test, the standard deviations (Std. Deviation) for the two groups were significantly different (0.260 and 0.499), indicating the tested variable groups had unequal variances. Thus results from the Equal Variances Not Assumed (EVNA) test were considered.

In the t-test for Equality of Means, the t statistic was -13.910, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 625.660. The probability from the t distribution with the stated degrees of freedom was indicated as 0.000 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was -0.382 and was the product of subtracting the sample mean for the second group (participants who used a support) from the sample mean for the first group (participants who did not use a support). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -0.436 to -0.328, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.000 Sig. (2-tailed) EVA ($p = 0.000$) and was less than 0.05, one can safely conclude the mean difference was not due to chance alone. Accompanying a report with a support containing analysis guidance in the form of footer, abstract, or interpretation guide has a significant, positive impact on the frequency of accurate conclusions educators draw concerning student achievement data when it is used.

An Independent Samples T-Test (see *Appendix I*) was also used to investigate the mere presence of an added support, regardless of whether or not participants reported using it. This test compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who received the

support and those who did not). As indicated in *Appendix I*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.000. This value was less than 0.10, suggesting the variable groups had unequal variances. Consistent with Levene's Test, the standard deviations (Std. Deviation) for the two groups were significantly different (0.318 and 0.453), indicating the tested variable groups had unequal variances. Thus results from the Equal Variances Not Assumed (EVNA) test were considered.

In the t-test for Equality of Means, the t statistic was -5.266, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 219.531. The probability from the t distribution with the stated degrees of freedom was indicated as 0.000 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was -0.175 and was the product of subtracting the sample mean for the second group (participants who received a support) from the sample mean for the first group (participants who did not receive a support). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -0.240 to -0.109, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.000 Sig. (2-tailed) ($p = 0.000$) and was less than 0.05, one can safely conclude the mean difference was not due to chance alone. Accompanying a report with a support containing analysis guidance in the form of footer,

abstract, or interpretation guide has a significant, positive impact on the frequency of accurate conclusions educators draw concerning student achievement data. In addition, this finding holds true whether or not the recipient indicates he or she uses the support.

Q2a. Research Question Q2a was asked as follows:

- What impact does a footer with analysis guidelines on a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that accompanying a report with a supportive footer containing analysis guidance would not have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.
- The alternative hypothesis was that accompanying a report with a supportive footer would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

The null hypothesis ($H2a_0$) was rejected and the alternative hypothesis was accepted ($H2a_a$) for Q2a based on the study results reported below. Accompanying a report with a supportive footer had a significant, positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

Table 4.02 features results shows for the 60 participants who received reporting environments containing footers. These 60 participants constituted 28% of the total 211-participant sample. These participants who had access to report footers indicated they used the footers 73% of the time. When these respondents had footers yet indicated they did not use the footers, their data analysis accuracy was 15%. All 60 participants with

footers, regardless of footer use, averaged a data analysis accuracy of 34%. In cases where respondents indicated they used the available footer, data analysis accuracy was 37%. In the 31 control group cases without any supports, which constituted 15% of the total 211-participant sample, data analysis accuracy was 11%. In terms of relative and absolute differences, educators' data analyses were 307% more accurate (with a 23 percentage point difference) when a footer was present and 336% more accurate (with a 26 percentage point difference) when respondents specifically indicated having used the footer (see *Figure 4.01*). See *Figure 4.03* for a visual representation of the footer's impact on educators' data analyses.

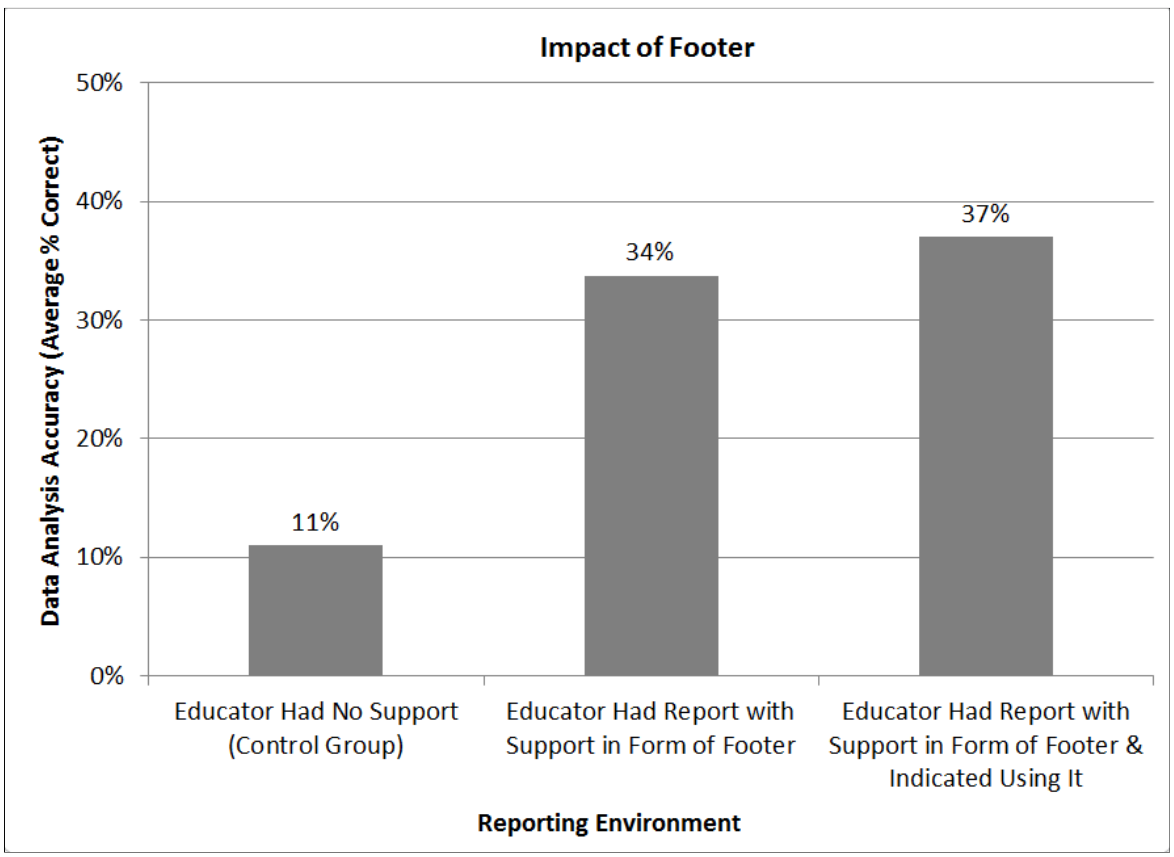


Figure 20: *Impact of Footer*

An Independent Samples T-Test (see *Appendix F*) was used to determine whether the footers' impact on educators' data analysis accuracy was significant. This test first compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who used the footer and those who did not). As indicated in *Appendix F*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.000. This value was less than 0.10, suggesting the variable groups had unequal variances. Consistent with Levene's Test, the standard deviations (Std. Deviation) for the two groups were significantly different (0.294 and 0.498), indicating the tested variable groups had unequal variances. Thus results from the Equal Variances Not Assumed (EVNA) test were considered.

In the t-test for Equality of Means, the t statistic was -8.022, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 275.119. The probability from the t distribution with the stated degrees of freedom was indicated as 0.000 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was -0.348 and was the product of subtracting the sample mean for the second group (participants who used the footer) from the sample mean for the first group (participants who did not use the footer). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -0.433 to -0.262, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.000 Sig. (2-tailed) ($p = 0.000$) and was less than 0.05, one can safely conclude the mean difference was not due to chance alone. Accompanying a report with a footer containing analysis guidance has a significant, positive impact on the frequency of accurate conclusions educators draw concerning student achievement data when it is used.

An Independent Samples T-Test (see *Appendix J*) was also used to investigate the mere presence of an added footer, regardless of whether or not participants reported using it. This test compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who received the footer and those who did not). As indicated in *Appendix J*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.000. This value was less than 0.10, suggesting the variable groups had unequal variances. The standard deviations (Std. Deviation) for the two groups were significantly different (0.318 and 0.474), indicating the tested variable groups had unequal variances. Thus results from the Equal Variances Not Assumed (EVNA) test were considered.

In the t-test for Equality of Means, the t statistic was -5.369, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 338.226. The probability from the t distribution with the stated degrees of freedom was indicated as 0.000 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was -0.225 and was the product of subtracting the sample mean for the second group (participants who received the footer) from the sample mean for the first group (participants who did not receive the footer). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -0.307 to -0.142, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.000 Sig. (2-tailed) ($p = 0.000$) and was less than 0.05, one can safely conclude the mean difference was not due to chance alone. Accompanying a report with a footer containing analysis guidance in the has a significant, positive impact on the frequency of accurate conclusions educators draw concerning student achievement data. In addition, this finding holds true whether or not the recipient indicates he or she uses the support.

Q2b. Research Question Q2b was asked as follows:

- What impact does the manner in which a footer is framed, in terms of moderate differences in length and text color, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that the manner in which a footer was framed, in terms of moderate differences in length and text color, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.

- The alternative hypothesis was that the manner in which a footer was framed, in terms of moderate differences in length and text color, would have an impact on the frequency of accurate conclusions educators drew concerning student achievement data.

The null hypothesis ($H2b_0$) was accepted and the alternative hypothesis was rejected ($H2b_a$) for Q2b based on the study results reported below. The manner in which a footer was framed, in terms of moderate differences in length and text color, did not have a significant impact on the frequency with which educators drew accurate conclusions concerning student achievement data. This is different than saying the manner in which a footer was framed did not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data. Rather, since it is already accepted the format of such tools *does* matter, generally-similar yet slightly-dissimilar footer formats were investigated in this study. See *Chapter 3: Research Method: Delimitations* for more details.

Table 4.02 features results shows for the 60 participants who received reporting environments containing footers, 30 of whom constituted 14% of the total 211-participant sample and received Footer A, and 30 of whom constituted 14% of the total 211-participant sample and received Footer B. Footer A was shorter and slightly less wordy (1st report footer: 39 words, 186 characters without spaces, 224 characters with spaces; 2nd report footer: 34 words, 156 characters without spaces, 228 characters with spaces) than the alternatively-framed footers and contained headings that utilized text color with meaning. Footer B was longer and slightly wordier (1st report footer: 58 words, 269 characters without spaces, 324 characters with spaces; 2nd report footer: 42 words, 199

characters without spaces, 237 characters with spaces) than the alternatively-framed footers and contained no headings or colored text.

Participants receiving Footer A indicated they used the footers 75% of the time, whereas participants receiving Footer B indicated they used the footers 70% of the time. When Footer A participants indicated they did not use the available footers, their data analysis accuracy was 27%, whereas when Footer B participants indicated they did not use the available footers, their data analysis accuracy was 6%. All 30 Footer A participants, regardless of footer use, averaged a data analysis accuracy of 36%, whereas all 30 Footer B participants, regardless of footer use, averaged a data analysis accuracy of 32%, In cases where respondents indicated they used the available footer, data analysis accuracy was 33% for Footer A participants and 40% for Footer B participants.

An Independent Samples T-Test (see *Appendix M*) was used to determine whether moderate changes in the footer's format, in terms of moderate differences in length and text color, had an impact on educators' data analysis accuracy that was significant. This test compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who received Footer A and those who received Footer B). As indicated in *Appendix M*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.803. This value was greater than 0.10, suggesting the variable groups had equal variances. In addition, the standard deviations (Std. Deviation) for the two groups were similar (32.618 and 33.434), indicating the tested variable groups had equal variances. Thus results from the Equal Variances Assumed (EVA) test were considered.

In the t-test for Equality of Means, the t statistic was 0.489, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 57.965. The probability from the t distribution with the stated degrees of freedom was indicated as 0.627 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was 4.167 and was the product of subtracting the sample mean for the second group (participants who received Footer A) from the sample mean for the first group (participants who received Footer B). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -12.904 to 21.237, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.627 Sig. (2-tailed) ($p = 0.627$) and was greater than 0.05, one can safely conclude the mean difference was due to chance alone. The manner in which a footer is framed, in terms of *moderate* differences in length and text color, does not have a significant impact on the frequency with which educators draw accurate conclusions concerning student achievement data.

Q3a. Research Question Q3a was asked as follows:

- What impact does providing a report abstract, such as a one-page reference sheet with report purpose and data use warnings specific to the report it accompanies, with a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that including a report abstract with a data system report would not have a positive impact on the frequency with which educators drew accurate conclusions concerning student achievement data.
- The alternative hypothesis was that including a report abstract with a report would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

The null hypothesis ($H3a_0$) was rejected and the alternative hypothesis was accepted ($H3a_a$) for Q3a based on the study results reported below. Including a report abstract with a report had a significant, positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

Table 4.02 features results shows for the 60 participants who received reporting environments containing abstracts. These 60 participants constituted 28% of the total 211-participant sample. These participants who had access to report abstracts indicated they used the abstracts 50% of the time. When these respondents had abstracts yet indicated they did not use the abstracts, their data analysis accuracy was 10%. All 60 participants with abstracts, regardless of abstract use, averaged a data analysis accuracy of 23%. In cases where respondents indicated they used the available abstract, data analysis accuracy was 33%. In the 31 control group cases without any supports, which constituted 15% of the total 211-participant sample, data analysis accuracy was 11%. In terms of relative and absolute differences, educators' data analyses were 205% more accurate (with a 12 percentage point difference) when an abstract was present and 300% more accurate (with a 22 percentage point difference) when respondents specifically

indicated having used the abstract (see *Figure 4.01*). See *Figure 4.04* for a visual representation of the abstract's impact on educators' data analyses.

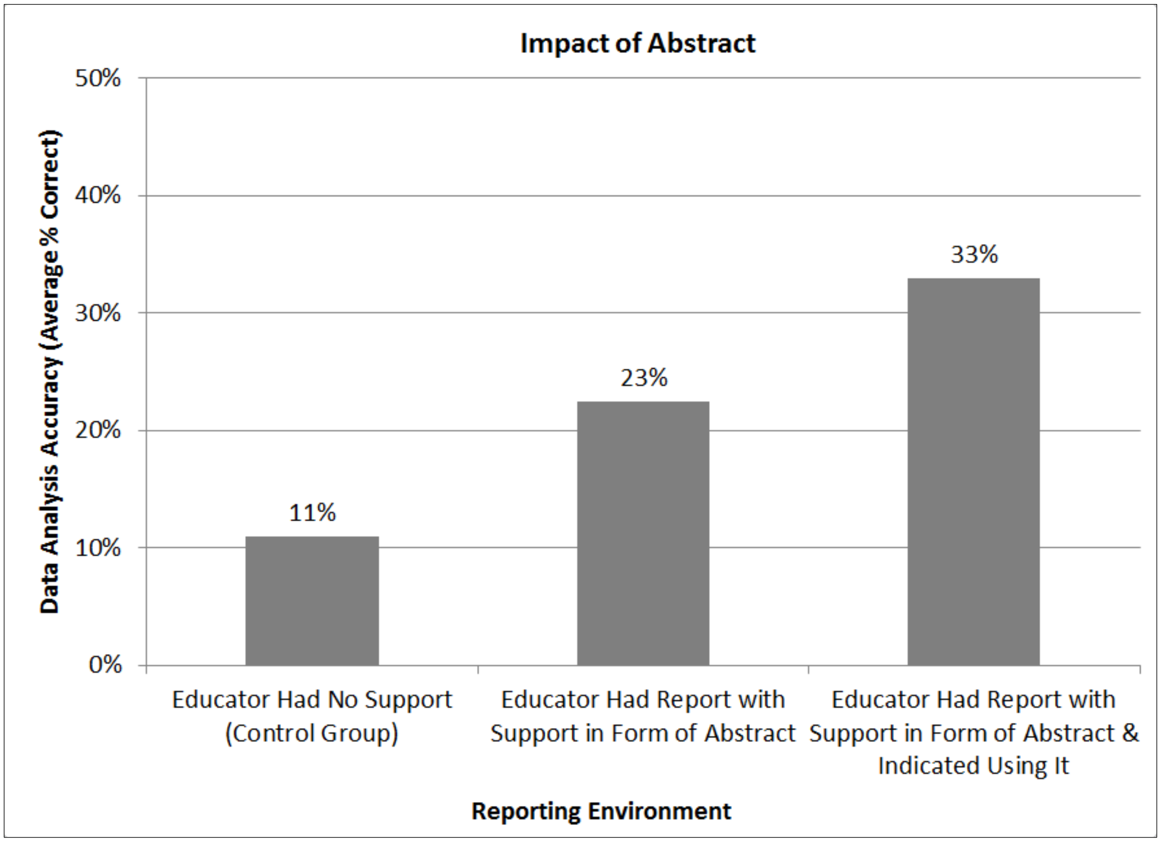


Figure 21: *Impact of Abstract*

An Independent Samples T-Test (see *Appendix G*) was used to determine whether the abstract' impact on educators' data analysis accuracy was significant. This test first compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who used the abstract and those who did not). As indicated in *Appendix G*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.000. This value was less than 0.10, suggesting the variable groups had unequal variances. Consistent with Levene's Test, the standard deviations (Std. Deviation) for the two groups were significantly different (0.298 and

0.484), indicating the tested variable groups had unequal variances. Thus results from the Equal Variances Not Assumed (EVNA) test were considered.

In the t-test for Equality of Means, the t statistic was -5.575, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 164.850. The probability from the t distribution with the stated degrees of freedom was indicated as 0.000 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was -0.268 and was the product of subtracting the sample mean for the second group (participants who used the abstract) from the sample mean for the first group (participants who did not use the abstract). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -0.363 to -0.173, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.000 Sig. (2-tailed) ($p = 0.000$) and was less than 0.05, one can safely conclude the mean difference was not due to chance alone. Accompanying a report with an abstract containing analysis guidance has a significant, positive impact on the frequency of accurate conclusions educators draw concerning student achievement data when it is used.

An Independent Samples T-Test (see *Appendix K*) was also used to investigate the mere presence of an added abstract, regardless of whether or not participants reported using it. This test compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who received the

abstract and those who did not). As indicated in *Appendix K*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.000. This value was less than 0.10, suggesting the variable groups had unequal variances. The standard deviations (Std. Deviation) for the two groups were significantly different (0.318 and 0.418), indicating the tested variable groups had unequal variances. Thus results from the Equal Variances Not Assumed (EVNA) test were considered.

In the t-test for Equality of Means, the t statistic was -2.853, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 312.890. The probability from the t distribution with the stated degrees of freedom was indicated as 0.005 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was -0.112 and was the product of subtracting the sample mean for the second group (participants who received the abstract) from the sample mean for the first group (participants who did not receiv the abstract). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -0.189 to -0.035, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.005 Sig. (2-tailed) ($p = 0.005$ to 0.009) and was less than 0.05, one can safely conclude the mean difference was not due to chance alone. Accompanying a report with an abstract containing analysis guidance has a significant, positive impact on the frequency of accurate conclusions educators draw

concerning student achievement data. In addition, this finding holds true whether or not the recipient indicates he or she uses the abstract.

Q3b. Research Question Q3b was asked as follows:

- What impact does the manner in which an abstract is framed, in terms of moderate differences in density and header color, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that the manner in which an abstract was framed, in terms of moderate differences in density and header color, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.
- The alternative hypothesis was that the manner in which an abstract was framed, in terms of moderate differences in density and header color, would have an impact on the frequency of accurate conclusions educators drew concerning student achievement data.

The null hypothesis ($H3b_0$) was accepted and the alternative hypothesis was rejected ($H3b_a$) for Q3b based on the study results reported below. The manner in which an abstract was framed, in terms of moderate differences in density and header color, did not have a significant impact on the frequency with which educators drew accurate conclusions concerning student achievement data. This is different than saying the manner in which an abstract was framed did not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data. Rather,

since it is already accepted the format of such tools *does* matter, generally-similar yet slightly-dissimilar abstract formats were investigated in this study. See *Chapter 3: Research Method: Delimitations* for more details.

Table 4.02 features results shows for the 60 participants who received reporting environments containing abstracts, 30 of whom constituted 14% of the total 211-participant sample and received Abstract A, and 30 of whom constituted 14% of the total 211-participant sample and received Abstract B. Abstract A was less dense and contained less information than the alternatively-framed abstracts and utilized heading color with meaning. Abstract B was denser and contained more information than the alternatively-framed abstracts and did not utilize heading color with meaning.

Participants receiving Abstract A indicated they used the abstracts 53% of the time, whereas participants receiving Abstract B indicated they used the abstracts 47% of the time. When Abstract A participants indicated they did not use the available abstracts, their data analysis accuracy was 11%, whereas when Abstract B participants indicated they did not use the available abstracts, their data analysis accuracy was 9%. All 30 Abstract A participants, regardless of abstract use, averaged a data analysis accuracy of 21%, whereas all 30 Abstract B participants, regardless of abstract use, averaged a data analysis accuracy of 24%, In cases where respondents indicated they used the available abstract, data analysis accuracy was 31% for Abstract A participants and 36% for Abstract B participants.

An Independent Samples T-Test (see *Appendix N*) was used to determine whether moderate changes in the abstract's format, in terms of moderate differences in density and header color, had an impact on educators' data analysis accuracy that was significant.

This test compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who received Abstract A and respondents who received Abstract B). As indicated in *Appendix N*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.365. This value was greater than 0.10, suggesting the variable groups had equal variances. Consistent with Levene's Test, the standard deviations (Std. Deviation) for the two groups were not significantly different (27.919 and 36.248), indicating the tested variable groups had equal variances, which was confirmed by an F-test ($F = .5932$, $p = .1657$). Thus results from the Equal Variances Assumed (EVA) test were considered.

In the t-test for Equality of Means, the t statistic was -0.399, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 58. The probability from the t distribution with the stated degrees of freedom was indicated as 0.691 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was -3.333 and was the product of subtracting the sample mean for the second group (participants who received Abstract A) from the sample mean for the first group (participants who received Abstract B). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -20.055 to 13.388, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.691 Sig. (2-tailed) ($p = 0.691$) and was greater than 0.05, one can safely conclude the mean difference was due to chance alone.

The manner in which an abstract is framed, in terms of *moderate* differences in density and header color, does not have an impact on the frequency with which educators draw accurate conclusions concerning student achievement data.

Q4a. Research Question Q4a was asked as follows:

- What impact does providing an interpretation guide, such as a two-sided reference sheet with analysis guidance and examples specific to the report it accompanies, with a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that including an interpretation guide with a data system report would not have a positive impact on the frequency with which educators drew accurate conclusions concerning student achievement data.
- The alternative hypothesis was that including an interpretation guide with a report would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

The null hypothesis ($H4a_0$) was rejected and the alternative hypothesis was accepted ($H4a_a$) for Q4a based on the study results reported below. Including an interpretation guide with a report had a significant, positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

Table 4.02 features results shows for the 60 participants who received reporting environments containing interpretation guides. These 60 participants constituted 28% of the total 211-participant sample. These participants who had access to report interpretation guides indicated they used the interpretation guides 52% of the time. When

these respondents had interpretation guides yet indicated they did not use the interpretation guides, their data analysis accuracy was 2%. All 60 participants with interpretation guides, regardless of interpretation guide use, averaged a data analysis accuracy of 30%. In cases where respondents indicated they used the available interpretation guide, data analysis accuracy was 48%. In the 31 control group cases without any supports, which constituted 15% of the total 211-participant sample, data analysis accuracy was 11%. In terms of relative and absolute differences, educators' data analyses were 273% more accurate (with a 19 percentage point difference) when an interpretation guide was present and 436% more accurate (with a 37 percentage point difference) when respondents specifically indicated having used the interpretation guide (see *Figure 4.01*). See *Figure 4.05* for a visual representation of the interpretation guide's impact on educators' data analyses.

An Independent Samples T-Test (see *Appendix H*) was used to determine whether the interpretation guide's impact on educators' data analysis accuracy was significant. This test first compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who used the interpretation guide and those who did not). As indicated in *Appendix H*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.000. This value was less than 0.10, suggesting the variable groups had unequal variances. Consistent with Levene's Test, the standard deviations (Std. Deviation) for the two groups were significantly different (0.257 and 0.499), indicating the tested variable groups had unequal variances. Thus results from the Equal Variances Not Assumed (EVNA) test were considered.

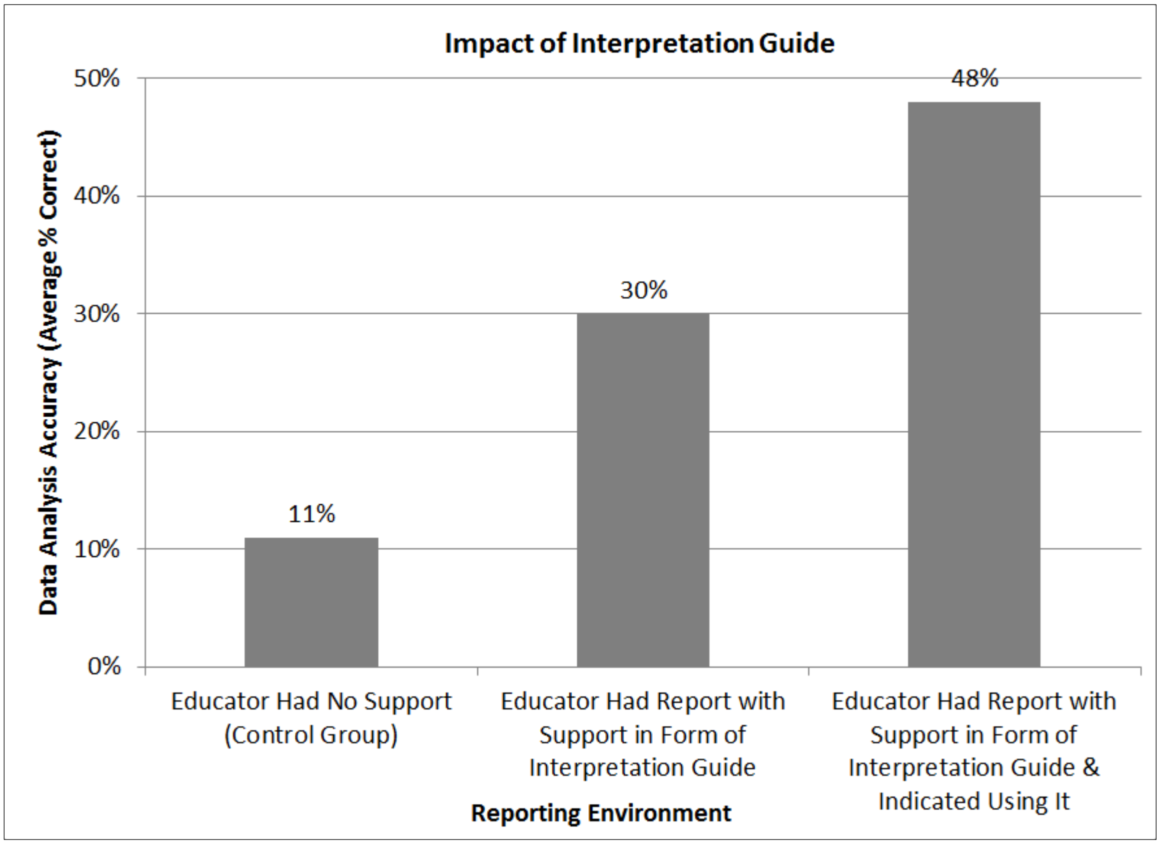


Figure 22: *Impact of Interpretation Guide*

In the t-test for Equality of Means, the t statistic was -10.166, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 157.550. The probability from the t distribution with the stated degrees of freedom was indicated as 0.000 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was -0.486 and was the product of subtracting the sample mean for the second group (participants who used the interpretation guide) from the

sample mean for the first group (participants who did not use the interpretation guide). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -0.580 to -0.391, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.000 Sig. (2-tailed) ($p = 0.000$) and was less than 0.05, one can safely conclude the mean difference was not due to chance alone. Accompanying a report with an interpretation guide containing analysis guidance has a significant, positive impact on the frequency of accurate conclusions educators draw concerning student achievement data when it is used.

An Independent Samples T-Test (see *Appendix L*) was also used to investigate the mere presence of an added interpretation guide, regardless of whether or not participants reported using it. This test compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who received the interpretation guide and those who did not). As indicated in *Appendix L*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.000. This value was less than 0.10, suggesting the variable groups had unequal variances. The standard deviations (Std. Deviation) for the two groups were significantly different (0.318 and 0.459), indicating the tested variable groups had unequal variances. Thus results from the Equal Variances Not Assumed (EVNA) test were considered.

In the t-test for Equality of Means, the t statistic was -4.547, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 332.451. The probability from the t distribution with the

stated degrees of freedom was indicated as 0.000 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was -0.187 and was the product of subtracting the sample mean for the second group (participants who received an interpretation guide) from the sample mean for the first group (participants who did not receive an interpretation guide). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -0.268 to -0.106 EVNA, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.000 Sig. (2-tailed) ($p = 0.000$) and was less than 0.05, one can safely conclude the mean difference was not due to chance alone. Accompanying a report with an interpretation guide containing analysis guidance has a significant, positive impact on the frequency of accurate conclusions educators draw concerning student achievement data. In addition, this finding holds true whether or not the recipient indicates he or she uses the interpretation guide.

Q4b. Research Question Q4b was asked as follows:

- What impact does the manner in which an interpretation guide is framed, in terms of moderate differences in length and information quantity, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that the manner in which an interpretation guide was framed, in terms of moderate differences in length and information quantity,

would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.

- The alternative hypothesis was that the manner in which an interpretation guide was framed, in terms of moderate differences in length and information quantity, would have an impact on the frequency of accurate conclusions educators drew concerning student achievement data.

The null hypothesis ($H4b_0$) was accepted and the alternative hypothesis was rejected ($H4b_a$) for Q4b based on the study results reported below. The manner in which an interpretation guide was framed, in terms of moderate differences in length and information quantity, did not have a significant impact on the frequency with which educators drew accurate conclusions concerning student achievement data. This is different than saying the manner in which an interpretation guide was framed did not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data. Rather, since it is already accepted the format of such tools *does* matter, generally-similar yet slightly-dissimilar interpretation guide formats were investigated in this study. See *Chapter 3: Research Method: Delimitations* for more details.

Table 4.02 features results shows for the 60 participants who received reporting environments containing interpretation guides, 30 of whom constituted 14% of the total 211-participant sample and received Interpretation Guide A, and 30 of whom constituted 14% of the total 211-participant sample and received Interpretation Guide B. Interpretation Guide A was shorter and contained less information (two pages) than the alternatively-framed interpretation guides and utilized heading color with meaning.

Interpretation Guide B was longer and slightly wordier (three pages) than the alternatively-framed interpretation guides and did not utilize heading color with meaning.

Participants receiving Interpretation Guide A indicated they used the interpretation guides 52% of the time, and participants receiving Interpretation Guide B also indicated they used the interpretation guides 52% of the time. When Interpretation guide A participants indicated they did not use the available interpretation guides, their data analysis accuracy was 0%, whereas when Interpretation Guide B participants indicated they did not use the available interpretation guides, their data analysis accuracy was 3%. All 30 Interpretation Guide A participants, regardless of interpretation guide use, averaged a data analysis accuracy of 32%, whereas all 30 Interpretation Guide B participants, regardless of interpretation guide use, averaged a data analysis accuracy of 28%, In cases where respondents indicated they used the available interpretation guide, data analysis accuracy was 48% for Interpretation Guide A participants and also 48% for Interpretation Guide B participants.

An Independent Samples T-Test (see *Appendix O*) was used to determine whether moderate changes in the interpretation guide's format, in terms of moderate differences in length and information quantity, had an impact on educators' data analysis accuracy that was significant. This test compared the means of a normally distributed interval dependent variable (analysis accuracy) for two independent groups (respondents who received Interpretation Guide A and those who received Interpretation Guide B). As indicated in *Appendix O*, the significance value (Sig.) of the Levene's Test for Equality of Variances statistic was 0.147. This value was greater than 0.10, suggesting the variable groups had equal variances. Consistent with Levene's Test, the standard deviations (Std.

Deviation) for the two groups were not significantly different (37.677 and 29.165), indicating the tested variable groups had equal variances, which was confirmed by an F-test ($F = 1.67$, $p = 0.17$). Thus results from the Equal Variances Assumed (EVA) test were considered.

In the t-test for Equality of Means, the t statistic was 0.383, which was calculated as the ratio of the difference between sample means divided by standard error of the difference. The total number of cases in both samples minus two, which was expressed as degrees of freedom (DF), was 58. The probability from the t distribution with the stated degrees of freedom was indicated as 0.703 Sig. (2-tailed); this was the probability of garnering an absolute value that was greater than or equal to the observed t statistic, if the difference between the sample means was considered purely random.

The mean difference was 3.333 and was the product of subtracting the sample mean for the second group (participants who received Interpretation Guide A) from the sample mean for the first group (participants who received Interpretation Guide B). The 95% Confidence Interval of the Difference that was used estimated the boundaries of -14.079 to 20.746, between which the true mean difference lay in 95% of all possible random samples of participants.

Since the p value, or Sig. (2-tailed), was 0.703 Sig. (2-tailed) ($p = 0.703$) and was greater than 0.05, one can safely conclude the mean difference was due to chance alone. The manner in which an interpretation guide is framed, in terms of *moderate* differences in length and information quantity, does not have a significant impact on the frequency with which educators draw accurate conclusions concerning student achievement data.

Q5a. Research Question Q5a was asked as follows:

- What impact does an educator's school site level type (i.e., elementary or secondary) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's school site level type (i.e., elementary or secondary) would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
- The alternative hypothesis was that an educator's school site level type (i.e., elementary or secondary) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H5a_0$) was rejected and the alternative hypothesis was accepted ($H5a_a$) for Q5a based on the study results reported below. An educator's school site level type (i.e., elementary or secondary) did not have a significant impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.04 features results for all 211 study participants, disaggregated by school level type. 132 participants, who constituted 63% of the total 211-participant sample, worked at the elementary school level. The elementary school level type typically begins with the starting grade level of transitional kindergarten (TK), preschool or pre-kindergarten (pre-K or PK), or kindergarten (K) and contains students up through grade 5 or 6. 79 participants, who constituted 37% of the total 211-participant sample, worked at the secondary school level. The secondary school level type typically begins with grade 5 or 6 and contains students up through grade 12. Elementary school respondents used supports 64% of the time, whereas secondary school respondents used supports 59% of

the time. All participants, regardless of whether or not they used supports, averaged 26% data analysis accuracy of at the elementary level and 27% data analysis accuracy of at the secondary level.

A crosstabulation table with Chi-square Test (see School Level Type section of *Appendix P*) was used to examine the relationship between the independent variable of school level type and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of school level type. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of school level type and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 4 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.550. The two-sided asymptotic significance of the Chi-square statistic for School Level Type was 0.538; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each school level type had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see School Level Type section of *Appendix Q*) was also used to examine the relationship between the independent variable of school level type and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of school level type. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of school level type and columns of support use were unrelated. The degrees of freedom (df), was 2 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.318. The two-sided asymptotic significance of the Chi-square statistic for School Level Type was 0.314; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each school level type had the same chance of using an analysis support. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship. An educator's school site level type (i.e., elementary or secondary) does not have a significant impact on the frequency of accurate conclusions he or she draws

concerning student achievement data. In addition, school level type does not have a significant impact on whether or not an educator uses an analysis support.

Q5b. Research Question Q5b was asked as follows:

- What impact does an educator's school site level (i.e., elementary, middle/junior high, or high school) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's school site level (i.e., elementary, middle/junior high, or high school) would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
- The alternative hypothesis was that an educator's school site level (i.e., elementary, middle/junior high, or high school) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H5b_0$) was rejected and the alternative hypothesis was accepted ($H5b_a$) for Q5b based on the study results reported below. An educator's school site level (i.e., elementary, middle/junior high, or high school) did not have a significant impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.05 features results for all 211 study participants, disaggregated by school level. 132 participants, who constituted 63% of the total 211-participant sample, worked at elementary schools. Like the elementary school level type, the elementary school level typically begins with the starting grade level of TK, PK, or K and contains students up

through grade 5 or 6. 47 participants, who constituted 22% of the total 211-participant sample, worked at middle schools or junior high schools. The middle/junior high school level type typically begins with grade 5 or 6 and contains students up through grade 8. 32 participants, who constituted 15% of the total 211-participant sample, worked at high schools. The high school level type typically begins with grade 9 and contains students up through grade 12. Elementary school respondents used supports 64% of the time, middle/junior high school respondents used supports 48% of the time, and high school respondents used supports 75% of the time. All participants, regardless of whether or not they used supports, averaged 26% data analysis accuracy of at the elementary school level, 25% data analysis accuracy of at the middle/junior high school level, and 30% data analysis accuracy of at the high school level.

A crosstabulation table with Chi-square Test (see School Level section of *Appendix P*) was used to examine the relationship between the independent variable of school level and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of school level. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of school level and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 8 for both the Pearson Chi-Square and

the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.730. The two-sided asymptotic significance of the Chi-square statistic for School Level was 0.551; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each school level had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see School Level section of *Appendix Q*) was also used to examine the relationship between the independent variable of school level and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of school level. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of school level and columns of support use were unrelated. The degrees of freedom (df), was 4 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.032. The two-sided asymptotic significance of the Chi-square statistic for School Level was 0.028; because this is less than 0.10 it is safe to conclude the differences are not due to

mere chance variation. This implies that each school level did not have the same chance of using an analysis support. An educator's school site level (i.e., elementary, middle/junior high, or high school) does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, school level has some impact on whether or not an educator uses an analysis support.

Q5c. Research Question Q5c was asked as follows:

- What impact does an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
- The alternative hypothesis was that an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H5c_0$) was rejected and the alternative hypothesis was accepted ($H5c_a$) for Q5c based on the study results reported below. An educator's school site academic performance, as measured by the 2012 Growth API, which is the California

state accountability measure, did not have a significant impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.06 features results for all 211 study participants, disaggregated by academic achievement of students at school sites, all of which were in California. The state of California's state accountability measure, which ranges from 200-1000 and is also used as a factor in federal accountability, is the Growth Academic Performance Index (API). Nine different 2012 Growth API scores were represented, ranging from 677 to 916, with a mean of 828. The API with the fewest participants was 916, which constituted 5% of the total 211-participant sample with 11 participants. The API with the most participants was 794, which constituted 16% of the total 211-participant sample with 33 participants.

The APIs where participants used supports 75% of the time, which was the most, were 677 and 893. The API where participants used supports 5% of the time, which was the least, was 916. The API with 41% data analysis accuracy, which was the most, was 827, whereas the API with 7% data analysis accuracy, which was the least, was 916. The API with participants who used supports more frequently tended to have higher data analysis accuracy. For example, the API with participants who used the least supports was also the API with the lowest data analysis accuracy.

A crosstabulation table with Chi-square Test (see Academic Performance section of *Appendix P*) was used to examine the relationship between the independent variable of API and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix*

P indicates the frequency of each data analysis accuracy score for each level of API. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of API and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 32 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.136. The two-sided asymptotic significance of the Chi-square statistic for Academic Performance was 0.397; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each API had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see Academic Performance section of *Appendix Q*) was also used to examine the relationship between the independent variable of API and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of API. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of API and columns of support use were unrelated. The degrees of freedom (df), was 16 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.018. The two-sided asymptotic significance of the Chi-square statistic for Academic Performance was 0.034; because this is less than 0.10 it is safe to conclude the differences are not due to mere chance variation. This implies that respondents of each API did not have the same chance of using an analysis support. An educator's school site academic performance, as measured by the 2012 Growth API, which is the California state accountability measure, does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, API has some impact on whether or not an educator uses an analysis support.

Q5d. Research Question Q5d was asked as follows:

- What impact does an educator's school site English Learner (EL) population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's school site English Learner (EL) population would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

- The alternative hypothesis was that an educator's school site English Learner (EL) population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H5d_0$) was rejected and the alternative hypothesis was accepted ($H5d_a$) for Q5d based on the study results reported below. An educator's school site EL population did not have a significant impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.07 features results for all 211 study participants, disaggregated by percent of the school site's students who are classified as English Learner (EL), sometimes also called English Language Learner (ELL). Nine different EL population levels were represented, ranging from 8% to 46%, with a mean of 29%. The EL population with the fewest participants was 16%, which constituted 5% of the total 211-participant sample with 11 participants. The EL population with the most participants was 30%, which constituted 16% of the total 211-participant sample with 33 participants.

The EL populations where participants used supports 75% of the time, which was the most, were 8% and 38%. The EL population where participants used supports 41% of the time, which was the least, was 16%. The EL population with 41% data analysis accuracy, which was the most, was 50, whereas the EL population with 7% data analysis accuracy, which was the least, was 16. The EL population with participants who used supports more frequently tended to have higher data analysis accuracy. For example, the EL population with participants who used the least supports was also the EL population with the lowest data analysis accuracy.

A crosstabulation table with Chi-square Test (see English Learner Population section of *Appendix P*) was used to examine the relationship between the independent variable of EL population and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of EL population. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of EL population and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 32 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.136. The two-sided asymptotic significance of the Chi-square statistic for English Learner Population was 0.397; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each EL population had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see English Learner Population section of *Appendix Q*) was also used to examine the relationship between the independent variable of EL population and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the

variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of EL population. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of EL population and columns of support use were unrelated. The degrees of freedom (df), was 16 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.018. The two-sided asymptotic significance of the Chi-square statistic for English Learner Population was 0.034; because this is less than 0.10 it is safe to conclude the differences are not due to mere chance variation. This implies that each EL population did not have the same chance of using an analysis support. An educator's school site EL population does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, EL population has some impact on whether or not an educator uses an analysis support.

Q5e. Research Question Q5e was asked as follows:

- What impact does an educator's school site Socioeconomically Disadvantaged population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's school site Socioeconomically Disadvantaged population would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
- The alternative hypothesis was that an educator's school site Socioeconomically Disadvantaged population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H5e_0$) was rejected and the alternative hypothesis was accepted ($H5e_a$) for Q5e based on the study results reported below. An educator's school site Socioeconomically Disadvantaged population did not have a significant impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.08 features results for all 211 study participants, disaggregated by percent of the school site's students who are classified as Socioeconomically Disadvantaged. Seven different socioeconomically disadvantaged population levels were represented, ranging from 22% to 78%, with a mean of 52%. The socioeconomically disadvantaged population with the fewest participants was 22%, which constituted 5% of the total 211-participant sample with 11 participants. The socioeconomically disadvantaged population with the most participants was 61%, which constituted 27% of the total 211-participant sample with 57 participants.

The socioeconomically disadvantaged population where participants used supports 75% of the time, which was the most, was 31%. The socioeconomically disadvantaged population where participants used supports 41% of the time, which was the least, was 22%. The socioeconomically disadvantaged population with 33% data analysis accuracy, which was the most, was 78%, whereas the socioeconomically

disadvantaged population with 7% data analysis accuracy, which was the least, was 22%. The socioeconomically disadvantaged population with participants who used supports more frequently tended to have higher data analysis accuracy. For example, the socioeconomically disadvantaged population with participants who used the least supports was also the socioeconomically disadvantaged population with the lowest data analysis accuracy.

A crosstabulation table with Chi-square Test (see Socioeconomically Disadvantaged Population section of *Appendix P*) was used to examine the relationship between the independent variable of Socioeconomically Disadvantaged population and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of Socioeconomically Disadvantaged population. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of Socioeconomically Disadvantaged population and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 24 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.140. The two-sided asymptotic significance of the Chi-square statistic for Socioeconomically Disadvantaged Population was 0.311; because this is greater than 0.10 it is safe to conclude the differences are due

to mere chance variation. This implies that each Socioeconomically Disadvantaged population had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see Socioeconomically Disadvantaged Population section of *Appendix Q*) was also used to examine the relationship between the independent variable of Socioeconomically Disadvantaged population and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of Socioeconomically Disadvantaged population. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of Socioeconomically Disadvantaged population and columns of support use were unrelated. The degrees of freedom (df), was 12 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.055. The two-sided asymptotic significance of the Chi-square statistic for Socioeconomically Disadvantaged Population was 0.091; because this is less than 0.10 it is safe to conclude the differences are not due to mere chance

variation. This implies that each Socioeconomically Disadvantaged population did not have the same chance of using an analysis support. An educator's school site Socioeconomically Disadvantaged population does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, Socioeconomically Disadvantaged population has some impact on whether or not an educator uses an analysis support.

Q5f. Research Question Q5f was asked as follows:

- What impact does an educators' school site Students with Disabilities population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's school site Students with Disabilities population would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
- The alternative hypothesis was that an educator's school site Students with Disabilities population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H5f_0$) was rejected and the alternative hypothesis was accepted ($H5f_a$) for Q5f based on the study results reported below. An educator's school site Students with Disabilities population did not have a significant impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.09 features results for all 211 study participants, disaggregated by percent of the school site's students who are classified as Students with Disabilities. Seven

different students with disabilities population levels were represented, ranging from 5% to 13%, with a mean of 10%. The students with disabilities population with the fewest participants was 5%, which constituted 8% of the total 211-participant sample with 16 participants. The students with disabilities population with the most participants was 9%, which constituted 18% of the total 211-participant sample with 38 participants.

The students with disabilities populations where participants used supports 75% of the time, which was the most, were 5% and 12%. The students with disabilities population where participants used supports 47% of the time, which was the least, was 11%. The students with disabilities populations with 31% data analysis accuracy, which was the most, were 9% and 13%, whereas the students with disabilities populations with 18% data analysis accuracy, which was the least, were 10% and 11%. The students with disabilities population with participants who used supports more frequently tended to have higher data analysis accuracy. For example, the students with disabilities population with participants who used the least supports was also the students with disabilities population with the lowest data analysis accuracy.

A crosstabulation table with Chi-square Test (see Students with Disabilities Population section of *Appendix P*) was used to examine the relationship between the independent variable of Students with Disabilities population and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of Students with Disabilities population. However,

from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of Students with Disabilities population and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 24 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.263. The two-sided asymptotic significance of the Chi-square statistic for Students with Disabilities Population was 0.530; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each Students with Disabilities population had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see Students with Disabilities Population section of *Appendix Q*) was also used to examine the relationship between the independent variable of Students with Disabilities population and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of Students

with Disabilities population. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of Students with Disabilities population and columns of support use were unrelated. The degrees of freedom (df), was 12 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.024. The two-sided asymptotic significance of the Chi-square statistic for Students with Disabilities Population was 0.043; because this is less than 0.10 it is safe to conclude the differences are not due to mere chance variation. This implies that each Students with Disabilities population did not have the same chance of using an analysis support. An educator's school site Students with Disabilities population does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, Students with Disabilities population has some impact on whether or not an educator uses an analysis support.

Q6a. Research Question 6a was asked as follows:

- What impact does an educator's veteran status have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's veteran status would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

- The alternative hypothesis was that an educator's veteran status would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H6a_0$) was rejected and the alternative hypothesis was accepted ($H6a_a$) for Q6a based on the study results reported below. An educator's veteran status did not have a significant impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.10 features results for all 211 study participants, disaggregated by veteran status in the form of how many years the participant had spent working as an educator, such as a teacher or administrator, for students under 19 years of age. Five different veteran statuses were represented: Less than 1 Year, Minimum of 5 Years, Minimum of 10 Years, Minimum of 15 Years, and Minimum of 20 Years. The veteran status of Less than 1 Year constituted 1% of the total 211-participant sample with 2 participants. The veteran status of Minimum of 5 Years constituted 9% of the total 211-participant sample with 20 participants. The veteran status of Minimum of 10 Years constituted 16% of the total 211-participant sample with 33 participants The veteran status of Minimum of 15 Years constituted 32% of the total 211-participant sample with 67 participants The veteran status of Minimum of 20 Years constituted 42% of the total 211-participant sample with 89 participants.

Participants who had been educators for less than one year used supports 75% of the time and averaged a data analysis accuracy of 25%. Participants who had been educators for at least five years used supports 70% of the time and averaged a data analysis accuracy of 35%. Participants who had been educators for at least 10 years used

supports 67% of the time and averaged a data analysis accuracy of 32%. Participants who had been educators for at least 15 years used supports 63% of the time and averaged a data analysis accuracy of 28%. Participants who had been educators for at least 20 years used supports 58% of the time and averaged a data analysis accuracy of 21%. The veteran status with participants who used supports more frequently tended to have higher data analysis accuracy. For example, the veteran status with participants who used the least supports was also the veteran status with the lowest data analysis accuracy.

A crosstabulation table with Chi-square Test (see Veteran Status section of *Appendix P*) was used to examine the relationship between the independent variable of veteran status and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of veteran status. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of veteran status and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 16 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.291. The two-sided asymptotic significance of the Chi-square statistic for Veteran Status was 0.393; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each veteran status had the

same chance of obtaining each data analysis accuracy score. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see Veteran status section of *Appendix Q*) was also used to examine the relationship between the independent variable of veteran status and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of veteran status. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of veteran status and columns of support use were unrelated. The degrees of freedom (df), was 8 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.279. The two-sided asymptotic significance of the Chi-square statistic for Veteran status was 0.336; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each veteran status had the same chance of using an analysis support. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship. An educator's veteran status does not have a significant impact on the frequency of accurate

conclusions he or she draws concerning student achievement data. In addition, veteran status has no significant impact on whether or not an educator uses an analysis support.

Q6b. Research Question 6b was asked as follows:

- What impact does an educator's current professional role (e.g., teacher, site/school administrator, etc.) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's current professional role (e.g., teacher, site/school administrator, etc.) would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
- The alternative hypothesis was that an educator's current professional role (e.g., teacher, site/school administrator, etc.) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H6b_0$) was rejected and the alternative hypothesis was accepted ($H6b_a$) for Q6b based on the study results reported below. An educator's current professional role (e.g., teacher, site/school administrator, etc.) did not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.11 features results for all 211 study participants, disaggregated by the educator's current professional role. Four different professional roles were represented: Teacher, Colleague Coach (e.g., Teacher on Special Assignment), Site/School Administrator, and District Administrator. The professional role of Teacher constituted

94% of the total 211-participant sample with 199 participants. The professional role of Colleague Coach constituted 1% of the total 211-participant sample with 2 participants. The professional role of Site/School Administrator constituted 4% of the total 211-participant sample with 8 participants. The professional role of District Administrator constituted 1% of the total 211-participant sample with 2 participants.

Teachers used supports 63% of the time and averaged a data analysis accuracy of 26%. Colleague coaches used supports 25% of the time and averaged a data analysis accuracy of 25%. Site/school administrators used supports 56% of the time and averaged a data analysis accuracy of 19%. District administrators used supports 100% of the time and averaged a data analysis accuracy of 75%. The professional role with participants who used supports more frequently tended to have higher data analysis accuracy. For example, the professional role with participants who used the most supports was also the professional role with the highest data analysis accuracy.

A crosstabulation table with Chi-square Test (see Role section of *Appendix P*) was used to examine the relationship between the independent variable of role and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of role. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one

could expect if the rows of role and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 12 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.417. The two-sided asymptotic significance of the Chi-square statistic for Role was 0.506; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each role had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see Role section of *Appendix Q*) was also used to examine the relationship between the independent variable of role and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of role. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of role and columns of support use were unrelated. The degrees of freedom (df), was 6 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.317. The

two-sided asymptotic significance of the Chi-square statistic for Role was 0.490; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each role had the same chance of using an analysis support. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship. An educator's current professional role (e.g., teacher, site/school administrator, etc.) does not have an impact on the frequency of accurate conclusions he or she draws concerning student achievement data. In addition, role has no significant impact on whether or not an educator uses an analysis support.

Q6c. Research Question 6c was asked as follows:

- What impact does an educator's perception of his or her own data analysis proficiency impact the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's perception of his or her own data analysis proficiency would be related to the frequency of accurate conclusions he or she drew concerning student achievement data.
- The alternative hypothesis was that an educator's perception of his or her own data analysis proficiency would not be related to the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H6c_0$) was rejected and the alternative hypothesis was accepted ($H6c_a$) for Q6c based on the study results reported below. An educator's perception of his

or her own data analysis proficiency was not related to the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.12 features results for all 211 study participants, disaggregated by perception of data analysis proficiency in the form of how participants rated their proficiency at analyzing student performance data. Four different perceived data analysis proficiency levels were represented: Very Proficient, Somewhat Proficient, Not Proficient, and Far from Proficient. Participants who rated themselves as Very Proficient constituted 21% of the total 211-participant sample with 45 participants. Participants who rated themselves as Somewhat Proficient constituted 66% of the total 211-participant sample with 139 participants. Participants who rated themselves as Not Proficient constituted 10% of the total 211-participant sample with 22 participants. Participants who rated themselves as Far from Proficient constituted 2% of the total 211-participant sample with 5 participants.

Participants who rated themselves as Very Proficient used supports 72% of the time and averaged a data analysis accuracy of 27%. Participants who rated themselves as Somewhat Proficient used supports 61% of the time and averaged a data analysis accuracy of 27%. Participants who rated themselves as Not Proficient used supports 57% of the time and averaged a data analysis accuracy of 23%. Participants who rated themselves as Far from Proficient used supports 30% of the time and averaged a data analysis accuracy of 10%. The perceived data analysis proficiency level with participants who used supports more frequently tended to have higher data analysis accuracy. For example, the perceived data analysis proficiency level with participants who used the least supports was also the perception of data analysis proficiency with the lowest data

analysis accuracy, and the perceived data analysis proficiency level with participants who used the most supports was also the perception of data analysis proficiency with the highest data analysis accuracy.

A crosstabulation table with Chi-square Test (see Perceived Data Analysis Proficiency section of *Appendix P*) was used to examine the relationship between the independent variable of perceived data analysis proficiency and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of perceived data analysis proficiency. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of perceived data analysis proficiency and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 12 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.901. The two-sided asymptotic significance of the Chi-square statistic for Perceived Data Analysis Proficiency was 0.950; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each perceived data analysis proficiency had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test

indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see Perceived Data Analysis Proficiency section of *Appendix Q*) was also used to examine the relationship between the independent variable of perceived data analysis proficiency and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of perceived data analysis proficiency. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of perceived data analysis proficiency and columns of support use were unrelated. The degrees of freedom (df), was 6 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.274. The two-sided asymptotic significance of the Chi-square statistic for Perceived Data Analysis Proficiency was 0.231; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each perceived data analysis proficiency had the same chance of using an analysis support. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship. An educator's

perception of his or her own data analysis proficiency is not related to the frequency of accurate conclusions he or she draws concerning student achievement data. In addition, perceived data analysis proficiency has no significant impact on whether or not an educator uses an analysis support.

Q6d. Research Question 6d was asked as follows:

- What impact does an educator's professional development over the past year, devoted specifically to *how* to analyze student data, have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's professional development over the past year, devoted specifically to how to analyze student data, would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
- The alternative hypothesis was that an educator's professional development over the past year, devoted specifically to how to analyze student data, would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H6d_0$) was rejected and the alternative hypothesis was accepted ($H6d_a$) for Q6d based on the study results reported below. An educator's professional development over the past year, devoted specifically to how to analyze student data, did not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.13 features results for all 211 study participants, disaggregated by data analysis professional development in the form of how many hours of PD the participant had taken part in within the past 12 months that specifically focused on learning how to correctly interpret student data. The related survey question respondents answered noted lots of professional development happens at school sites – for example, demonstrations to accompany textbook adoptions, meetings with colleagues to share differentiation strategies, training on how to use new software, etc. – yet only some professional development specifically focuses on how to analyze student data. Different amounts of data analysis professional development were represented: 0 Hours, Minimum of 1 Hour, Minimum of 2 Hours, Minimum of 5 Hours, and Minimum of 8 Hours. Participants who had undergone 0 Hours of data analysis PD in the last year constituted 41% of the total 211-participant sample with 87 participants. Participants who had undergone Minimum of 1 Hour of data analysis PD in the last year constituted 23% of the total 211-participant sample with 48 participants. Participants who had undergone Minimum of 2 Hours of data analysis PD in the last year constituted 18% of the total 211-participant sample with 39 participants. Participants who had undergone Minimum of 5 Hours of data analysis PD in the last year constituted 9% of the total 211-participant sample with 19 participants. Participants who had undergone Minimum of 8 Hours of data analysis PD in the last year constituted 9% of the total 211-participant sample with 18 participants.

Participants who had undergone 0 Hours of data analysis PD used supports 58% of the time and averaged a data analysis accuracy of 23%. Participants who had undergone Minimum of 1 Hour of data analysis PD used supports 63% of the time and averaged a data analysis accuracy of 26%. Participants who had undergone Minimum of

2 Hours of data analysis PD used supports 72% of the time and averaged a data analysis accuracy of 30%. Participants who had undergone Minimum of 5 Hours of data analysis PD used supports 71% of the time and averaged a data analysis accuracy of 22%. Participants who had undergone Minimum of 8 Hours of data analysis PD used supports 53% of the time and averaged a data analysis accuracy of 36%.

A crosstabulation table with Chi-square Test (see Professional Development (PD) section of *Appendix P*) was used to examine the relationship between the independent variable of PD and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of PD. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of PD and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 16 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.754. The two-sided asymptotic significance of the Chi-square statistic for Professional Development (PD) was 0.713; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each PD had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test

indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see Professional Development (PD) section of *Appendix Q*) was also used to examine the relationship between the independent variable of PD and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of PD. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of PD and columns of support use were unrelated. The degrees of freedom (df), was 8 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.149. The two-sided asymptotic significance of the Chi-square statistic for Professional Development (PD) was 0.185; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each PD had the same chance of using an analysis support. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship. An educator's professional development over the past year, devoted specifically to how to analyze student data, does not have an impact on the frequency of

accurate conclusions he or she draws concerning student achievement data. In addition, PD has no significant impact on whether or not an educator uses an analysis support.

Q6e. Research Question 6e was asked as follows:

- What impact does the number of graduate-level educational measurement courses an educator has taken have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null and alternative hypotheses for this question were, respectively:

- The null hypothesis was that an educator's number of graduate-level educational measurement courses would have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.
- The alternative hypothesis was that an educator's number of graduate-level educational measurement courses would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

The null hypothesis ($H6e_0$) was rejected and the alternative hypothesis was accepted ($H6e_a$) for Q6e based on the study results reported below. An educator's number of graduate-level educational measurement courses did not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.14 features results for all 211 study participants, disaggregated by educational measurement course number in the form of how many graduate-level courses the participant had taken that were specifically dedicated to educational measurement. The related survey question respondents answered noted *educational measurement* refers to the analysis of student assessment data to draw conclusions about abilities, and graduate-level courses specifically dedicated to educational measurement might include

such topics as student performance data analysis, measurement theory, or psychometrics. Five different educational measurement course numbers were represented: 0 Courses, Minimum of Course, Minimum of 2 Courses, Minimum of 3 Courses, and Minimum of 4 Courses. Participants who had taken 0 educational measurement courses constituted 47% of the total 211-participant sample with 100 participants. Participants who had taken at least 1 educational measurement course constituted 24% of the total 211-participant sample with 51 participants. Participants who had taken at least 2 educational measurement courses constituted 17% of the total 211-participant sample with 35 participants. Participants who had taken at least 3 educational measurement courses constituted 5% of the total 211-participant sample with 11 participants. Participants who had taken at least 4 educational measurement courses constituted 7% of the total 211-participant sample with 14 participants.

Participants who had taken 0 educational measurement courses used supports 55% of the time and averaged a data analysis accuracy of 23%. Participants who had taken at least 1 educational measurement course used supports 70% of the time and averaged a data analysis accuracy of 30%. Participants who had taken at least 2 educational measurement courses used supports 73% of the time and averaged a data analysis accuracy of 29%. Participants who had taken at least 3 educational measurement courses used supports 64% of the time and averaged a data analysis accuracy of 25%. Participants who had taken at least 4 educational measurement courses used supports 61% of the time and averaged a data analysis accuracy of 27%. The educational measurement course number with participants who used supports more frequently tended to have higher data analysis accuracy. For example, the educational measurement course

number with participants who used the least supports was also the educational measurement course number with the lowest data analysis accuracy.

A crosstabulation table with Chi-square Test (see Graduate Educational Measurement Courses section of *Appendix P*) was used to examine the relationship between the independent variable of graduate educational measurement courses and data analysis accuracy. This approach was also used to identify whether the variable had a significant impact on educators' data analysis accuracy that might be of import to the study's primary research questions. The related *Count* section of *Appendix P* indicates the frequency of each data analysis accuracy score for each level of graduate educational measurement courses. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix P* and what one could expect if the rows of graduate educational measurement courses and columns of data analysis accuracy scores were unrelated. The degrees of freedom (df), was 16 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.548. The two-sided asymptotic significance of the Chi-square statistic for Graduate Educational Measurement Courses was 0.677; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each graduate educational measurement courses had the same chance of obtaining each data analysis accuracy score. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship.

A crosstabulation table with Chi-square Test (see Graduate Educational Measurement Courses section of *Appendix Q*) was also used to examine the relationship between the independent variable of graduate educational measurement courses and the educator's likelihood of using analysis supports when they were available. This approach was also used to identify whether the variable had a significant impact on educators' likelihood of using a support, as such information could have been of import to the study's primary research questions. The related *Count* section of *Appendix Q* indicates the frequency with which analysis supports were used or wanted by respondents of each level of graduate educational measurement courses. However, from the crosstabulation table alone one cannot conclude whether these differences are real or merely due to chance variation.

Thus the Pearson Chi-square test was conducted to measure the discrepancy between the cell counts shown in the related *Count* section of *Appendix Q* and what one could expect if the rows of graduate educational measurement courses and columns of support use were unrelated. The degrees of freedom (df), was 8 for both the Pearson Chi-Square and the Likelihood Ratio, which had a two-sided asymptotic significance, shown as *Asymp. Sig. (2-sided)*, of 0.336. The two-sided asymptotic significance of the Chi-square statistic for Graduate Educational Measurement Courses was 0.338; because this is greater than 0.10 it is safe to conclude the differences are due to mere chance variation. This implies that each graduate educational measurement courses had the same chance of using an analysis support. Since the Chi-square test indicated no relationship, no additional symmetric measures were necessary to indicate such the strength of a relationship. An educator's number of graduate-level educational measurement courses

does not have an impact on the frequency of accurate conclusions he or she draws concerning student achievement data. In addition, graduate educational measurement courses has no significant impact on whether or not an educator uses an analysis support.

Evaluation of Findings

All supports used in the study – footers, abstracts, and interpretation guides – had a significant, positive impact on the participating educators' data analysis accuracy. This resulted in acceptance of the alternative hypotheses for primary Research Questions Q1, Q2a, Q3a, and Q4a. Specifically, in terms of relative and absolute differences, educators' data analyses were:

- 264% more accurate (with an 18 percentage point difference) when any one of the three supports was present and 355% more accurate (with a 28 percentage point difference) when respondents specifically indicated having used the support,
- 307% more accurate (with a 23 percentage point difference) when a footer was present and 336% more accurate (with a 26 percentage point difference) when respondents specifically indicated having used the footer,
- 205% more accurate (with a 12 percentage point difference) when an abstract was present and 300% more accurate (with a 22 percentage point difference) when respondents specifically indicated having used the abstract, and
- 273% more accurate (with a 19 percentage point difference) when an interpretation guide was present and 436% more accurate (with a 37 percentage point difference) when respondents specifically indicated having used the interpretation guide.

The results were expected to be positive *when supports were used* given previously-existing literature recommending the presence of footers, abstracts, and interpretation guides. However, some literature suggested the supports would not be utilized and would be rendered ineffective. Not only did the supports prove to have a significant, positive impact on data analysis accuracy, but the substantial rate at which they were utilized rendered their value significant for *all* educators as a whole, even when respondents' use of the supports was not considered. Nonetheless, respondents' data analyses were even higher when they indicated having used the available support.

The minor modifications in support format, mainly in terms of length and color usage, had no significant impact on the participating educators' data analysis accuracy. This resulted in acceptance of the null hypotheses for primary Research Questions Q2b, Q3b, and Q4b. These results were somewhat unexpected given literature on behavioral economics, particularly in the area of framing, and literature on report and documentation design. However, it is important to note all support format variations used in the study subscribed to best practices recommended in literature on report and documentation design. Thus the variations were minor and designed to garner more specificity in these best practices. It was thus concluded such minor variations are also minor in their impact on educators' data analyses.

Additional, secondary research questions were used to add insight to the primary research questions. Findings in relation to these questions determined that educators' school site demographics had no significant impact on their data analysis accuracy that might impact the primary research questions. In other words, an educator's school level type, school level, academic performance, EL population, Socioeconomically

Disadvantaged population, or Students with Disabilities population had no significant impact on data analysis accuracy. This resulted in acceptance of the alternative hypotheses for secondary Research Questions Q5a, Q5b, Q5c, Q5d, Q5e, and Q5f. These results were expected given the lack of literature indicating the impact of such school site demographic variables. These variables were examined, nonetheless, given common-yet-unsubstantiated theories they are of import to data analyses and thus to support use and effectiveness.

Likewise, findings in relation to the secondary questions determined that educators' demographics had no significant impact on their data analysis accuracy that might impact the primary research questions. In other words, an educator's veteran status, current professional role, perception of his or her own data analysis proficiency, data analysis PD time, and number of graduate-level educational measurement courses had no significant impact on data analysis accuracy. This resulted in acceptance of the alternative hypotheses for secondary Research Questions Q6a, Q6b, Q6c, Q6d, and Q6e. These results were expected given the lack of literature indicating the impact of such educator demographic variables. These variables were examined, nonetheless, given common-yet-unsubstantiated theories they are of import to data analyses and thus to support use and effectiveness.

Summary

Data-informed decisions can improve learning (Sabbah, 2011; Underwood, Zapata-Rivera, & VanWinkle, 2010; Wohlstetter, Datnow, & Park, 2008), yet this requires decisions to be data-*informed* rather than data-*mis*informed. Unfortunately, there is clear evidence many users of data system reports have trouble understanding the data

(Hattie, 2010; National Research Council, 2001; Wayman et al., 2010; Zwick et al., 2008). For example, in a national study of districts known for *strong* data use, teachers only correctly interpreted 48% of data (U.S. Department of Education Office of Planning, Evaluation and Policy Development [USDEOPEPD], 2009). Even district-level educators find student data system reports to be complex, hard to read, and even harder to interpret (Underwood, Zapata-Rivera, & VanWinkle, 2010). Yet labeling and tools within data systems to assist analysis are uncommon (USDEOPEPD, 2009).

The *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study was used to determine the degree to which three forms of data system-embedded data analysis support can improve the accuracy of educators' data analyses:

- (a) labeling in the form of brief, cautionary verbiage in report footers; and
- (b) supplemental documentation in the form of report abstracts and
- (c) interpretation guides.

All supports used in the study had a significant, positive impact on the participating educators' data analysis accuracy, and this relationship held true even when recipients did not indicate they used the supports.

Although two differently-framed forms of each of these supports was tested, there was no significant difference in data analysis accuracy rendered. However, the framing differences were slight and thus this finding should not be mistaken as an indication that framing does not matter. The impact of educators' school site demographics and personal demographics was also explored in relation to secondary research questions in the event they proved to have a significant impact on data analysis accuracy that should be considered. However, none of these variables had a significant impact on the accuracy of

participants' data analyses. Thus the findings concerning the effectiveness of report footers, abstracts, and interpretation guides apply equally to educators of varied demographics and varied school site demographics.

Chapter 5: Implications, Recommendations, and Conclusions

Before the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study, it was undecided whether adding specific over-the-counter data supports to data systems can reduce the number of analysis errors, and to what degree. Educators worldwide test students, distribute score reports, and expect stakeholders to make improvements based on these reports (Hattie & Brown, 2008). Most educators have access to data systems to generate and analyze score reports (Aarons, 2009; Herbert, 2011). Yet educators do not use this data correctly, and there is clear evidence many users of data system reports have trouble understanding the data (Hattie, 2010; National Research Council, 2001; Wayman et al., 2010; Zwick et al., 2008).

Data use impacts students, and misunderstandings when using data systems can cripple data use in school districts (Wayman, Cho, & Shaw, 2009). Despite this, labeling and tools within data systems to assist analysis are uncommon (USDEOPEPD, 2009). There is a clear need for research identifying how reports can better facilitate correct interpretations by its users (Goodman & Hambleton, 2004; Hattie, 2010).

The *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study rendered findings that data system-embedded data analysis support in the forms of footers, abstracts, and interpretation guides all have a positive, significant impact on the accuracy of educators' data analyses. The experimental quantitative study was used to measure educators' data analysis accuracy when using typical data system reports, which do not contain analysis guidance on the reports or by way of supplemental documentation. Results from that control group were compared to those of educators analyzing data in reports containing analysis guidance in the form of (a) labeling directly

on the report through a footer, or in the form of (b) supplemental documentation through an abstract or (c) interpretation guide. The research design also allowed for framing influences by presenting each of the three data analysis supports (a-c) in two moderately-different formats. This allowed the study to measure not only whether – and to what extent – each analysis support can increase analysis accuracy, but also the more effective way in which to frame each support. The impact of these moderate format changes were found to be insignificant.

The study dealt exclusively with educators and their use of data system reports and resources in an isolated setting. Thus, to maintain external validity, study findings may not be applied to inferences concerning non-educators, such as parents, students, or politicians. Likewise, in consideration of the potential impact of interaction of setting and treatment, no generalizations of data analyses may be made of analysis environments that are not report-based, such as data analyses made based on data group discussions or based on an explanation heard by a data coach.

Deliberate measures were taken to ensure the study adhered to ethical practices, such as protection from harm, informed consent, right to privacy, and honesty with professional colleagues. The researcher took key steps at each stage of the doctoral process to apply the care and integrity needed to meet the ethical standards of scientific research. These steps encompassed considerations such as preventing plagiarism; risk assessment; informed consent; privacy, confidentiality, and data handling; protecting the integrity of results reporting and its ability to be applied to real world practice; overcoming threats to construct, external, and internal validity; awareness of procedures for mistakes and negligence, and piori IRB approval.

This chapter contains implications of the study's findings, which are organized around the study's research questions. Next the chapter contains recommendations for practical applications of the study findings, as well as recommendations for future research. The chapter's key implications and recommendations are then summarized.

Implications

In this section the results communicated in *Chapter 4: Findings* are explained in terms of their implications. When this section contains reference to:

- *supports*, it is referring to (a) any support, combining the supports that follow as b-d; (b) footer; (c) abstract; or (d) interpretation guide.
- *support use*, it is referring to instances in which respondents indicated they (a) used the available support or (b) would have used a support, as was a response option for control group participants who did not receive any supports. Note the support use refers to a percent of *instances* and not a percent of *participants*, as a single respondent might have used supports in only a portion of the instances to which he or she was exposed to the support, such as using the footer on Report 1 but not the footer on Report 2.
- *data analysis accuracy*, it is referring to the mean value of participants' percent correct scores earned when answering Questions 4-7 measuring data analysis accuracy.

The results featured in *Tables 4.01-4.14*, which are organized around the study's research questions, can serve as helpful references while reading about the results' implications. Research Questions were comprised of Q1-Q3b, which constituted the study's seven primary research questions, and Q4a-Q6e, which constituted the study's 11 secondary

research questions serving the sole role of informing implications addressed by the primary research questions.

Research Questions Q1, Q2a, Q3a, and Q4a. Research Questions Q1, Q2a, Q3a, and Q4a were all answered with significant findings concerning the value of the varied supports they concerned. In summary, these questions and their accepted hypotheses are featured below with question-specific implications, followed by implications that relate to all four research questions.

Q1. Research Question Q1 was asked as follows:

- What impact does data analysis guidance accompanying a data system report in the form of footer, abstract, or interpretation guide have on how frequently educators draw accurate conclusions concerning student achievement data?

The null hypothesis ($H1_0$) was rejected and the following alternative hypothesis was accepted ($H1_a$) for Q1 based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that accompanying a report with a support containing analysis guidance in the form of footer, abstract, or interpretation guide would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

In terms of relative and absolute differences, educators' data analyses were 264% more accurate (with an 18 percentage point difference) when any one of the three supports was present and 355% more accurate (with a 28 percentage point difference) when respondents specifically indicated having used the support. These findings imply there are direct benefits to educators' data use when a data system and its reports embed any

one of the three data analysis supports investigated in this study. More implications will be explained later in this *Research Questions Q1, Q2a, Q3a, and Q4a* section.

Q2a. Research Question Q2a was asked as follows:

- What impact does a footer with analysis guidelines on a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?

The null hypothesis ($H2a_0$) was rejected and the following alternative hypothesis was accepted ($H2a_a$) for Q2a based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that accompanying a report with a supportive footer would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

In terms of relative and absolute differences, educators' data analyses were 307% more accurate (with a 23 percentage point difference) when a footer was present and 336% more accurate (with a 26 percentage point difference) when respondents specifically indicated having used the footer. These findings imply there are direct benefits to educators' data use when data reports include a footer offering data analysis support. More implications will be explained later in this *Research Questions Q1, Q2a, Q3a, and Q4a* section.

Q3a. Research Question Q3a was asked as follows:

- What impact does providing a report abstract, such as a one-page reference sheet with report purpose and data use warnings specific to the report it accompanies,

with a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?

The null hypothesis ($H3a_0$) was rejected and the following alternative hypothesis was accepted ($H3a_a$) for Q3a based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that including a report abstract with a report would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

In terms of relative and absolute differences, educators' data analyses were 205% more accurate (with a 12 percentage point difference) when an abstract was present and 300% more accurate (with a 22 percentage point difference) when respondents specifically indicated having used the abstract. These findings imply there are direct benefits to educators' data use when data systems offer report-specific abstracts offering data analysis support. More implications will be explained later in this *Research Questions Q1, Q2a, Q3a, and Q4a* section.

Q4a. Research Question Q4a was asked as follows:

- What impact does providing an interpretation guide, such as a two-sided reference sheet with analysis guidance and examples specific to the report it accompanies, with a data system report have on how frequently educators draw accurate conclusions concerning student achievement data?

The null hypothesis ($H4a_0$) was rejected and the following alternative hypothesis was accepted ($H4a_a$) for Q4a based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that including an interpretation guide with a report would have a positive impact on the frequency of accurate conclusions educators drew concerning student achievement data.

In terms of relative and absolute differences, educators' data analyses were 273% more accurate (with a 19 percentage point difference) when an interpretation guide was present and 436% more accurate (with a 37 percentage point difference) when respondents specifically indicated having used the interpretation guide. These findings imply there are direct benefits to educators' data use when data systems offer report-specific interpretation guides offering data analysis support. More implications will be explained below.

Educators want data system/report-embedded supports. 87% of control group participants, who did not receive any supports, indicated they would have used the added support if they had it in the form of a footer, abstract, or interpretation guide (see *Table 4.02*). This finding supported experts' assertions that educators desire more data analysis support from their data systems and its reports. For example, teachers expressed a need for easier ways to use data, are overwhelmed by data, and have to work longer hours to use data (Wayman et al., 2010). Because of the now-proven support footers, abstracts, and interpretation guides provide, it has been shown these resources can help to fulfill the need educators' expressed. The implication that educators want data system/report-embedded data analysis supports was further substantiated by the results for the 180 participants who received reporting environments containing supports, as these participants who had access to report supports indicated they used the supports 58% of the time (see *Table 4.02*).

Educators struggle with data analyses. Control group participants, who received no added supports, averaged a data analysis accuracy of 11% correct, which was grossly below the scores of educators in supported environments noted earlier. This finding supports field literature assertions that educators using typical data system reports struggle to make accurate data analyses. For example, many teachers and administrators do not know fundamental analysis concepts, and educators are not skilled at using data daily to improve student learning, which is a needed skill in educator professions (Zwick et al., 2008). Not all educators have the skills needed to successfully use data to inform decisions, and having data does not mean it will be used properly (Marsh et al., 2006). Few educators automatically know how to use available data effectively (DQC, 2009).

Many educators experience difficulties just trying to understand the data they are analyzing (Goodman & Hambleton, 2004; Hambleton, 2002; Hattie, 2010; NRC, 2001). For example, teachers have difficulty using data systems to interpret student data, even amongst teachers who serve as assessment coaches to their peers (Underwood et al., 2008). The problem is not restricted to teachers. Stakeholders at all levels have trouble interpreting data, such as principals who are intimidated by data and need training, and teacher coaches who are not tech-savvy and have trouble sharing assessments and data system knowledge with teachers (Underwood et al., 2008). State-level stakeholders are also at varying stages of being able to actually analyze the data that data systems display (Minnici & Hill, 2007). Even at the state level, stakeholders are not using student data effectively (Halpin & Cauthen, 2011).

Although the three supports investigated in this study increased educators' data analysis accuracy by 205%-307% when they were merely present, and by even more –

300%-436% – when participants specifically indicated they used them (see *Figure 4.01*), no single support resulted in 100% data analysis accuracy for *all* users. The average data analysis accuracy only rose to up to 48% correct (see *Table 4.02*). See *Chapter 5: Implications, Recommendations, and Conclusions: Recommendations: All three supports simultaneously* for related recommendations.

Support benefits persist regardless of report or question type. *Table 4.03* features results by data analysis question on the survey in order to address any questions about whether there was an imbalance in the questions used to measure the data analysis accuracy pertinent to all of the study's research questions, particularly Q1, Q2a, Q3a, and Q4a. As *Table 4.03* shows, there was little difference between participants' data analysis accuracy on each question and on each support:

- Participants' data analysis accuracy was 29% on Question 4 and 28% on Question 5, with an average data analysis accuracy of 28% for Report 1 questions.
- Participants' data analysis accuracy was 21% on Question 6 and 27% on Question 7, with an average data analysis accuracy of 24% for Report2 questions.

Disaggregating results by question and report was important in order to determine if there were any report-type or question-type limitations to the data analysis accuracy impact measured in relation to all of the study's research questions, particularly Q1, Q2a, Q3a, and Q4a. Report differences, to which all 211 participants were exposed, included:

- Report 1 was graphical in format, whereas Report 2 was tabular in format.
- Report 1 utilized the use of a key/legend to answer analysis questions, whereas Report 2 did not.

- Color was vital to the understanding of Report 1 data, whereas color was not pertinent to the analysis of Report 2 data.
- Report 1 related to an assessment considered higher stakes than the Report 2 assessment.
- Report 1 presented aggregate data in the form of site and state averages, whereas Report 2 presented student-level data.

Question differences, to which all 211 participants were exposed, included:

- Questions 4-5 analyses required more steps than Questions 6-7 analyses, presenting varied levels of critical thinking and difficulty.
- Questions 4-5 each required the selection of only one of the multiple-choice answer options, whereas Questions 6-7 each required the selection of one or more of the multiple-choice answer options, with the correct number of selections that must be made left as undefined for respondents as the correct answers.

While this variety resulted in increased triangulation for the study, it also contributed to the implication that the study's three analysis supports proved effective when used with any of the report types and in answering any of the question types. This implication is supported by the support success findings noted earlier, combined with the fact that there were insignificant difference in educators' data analysis accuracy question-to-question and report-to-report. For example, respondents averaged 29% data analysis accuracy on Question 4, 28% data analysis accuracy on Question 5, 21% data analysis accuracy on Question 6, and 27% data analysis accuracy on Question 7. Likewise, respondents averaged 28% for Report 1 questions and 24% data analysis accuracy for Report2 questions.

Data-informed decision-making and helping students. As explained in *Chapter 4: Findings: Results*, all three data analysis supports used in the study were found to be significantly beneficial to educators' data analyses. This finding holds implications for data-informed decision-making, often called data-driven decision-making, of which data analysis is a key step. Research review indicates using data to inform instructional decisions can result in greater student achievement (Lewis, Madison-Harris, Muoneke, Times, 2010; Wayman, 2005; Wohlstetter et al., 2008). Thus educators realize data can be the foundation for action toward school improvement (Sabbah, 2011; Supovtiz & Klein, 2003). Worldwide, nations and U.S. states use some form of national or state-wide testing; distribute score reports to students, parents, educators, and/or government; and expect stakeholders to learn from these reports and use them for data-informed decision-making (Hattie & Brown, 2008). However, If data system users do not understand how to properly analyze data, data used will be used incorrectly (NFES, 2011).

Even the name of the premise these stakeholders are employing – data-*informed* decision-making – indicates it relies on the understanding that the data is being used to *inform* decisions, not *mis*inform them. Misunderstandings about how to use data and a data system can cripple data use in a school district and cause low data system use rates and resistance to data (Wayman et al., 2009). Conversely, if used correctly, data use can lead to insight into students' abilities and to decisions to improve instruction (Underwood et al., 2010). Since this study resulted in confirmation that three, specific supports in data systems and their reports improve educators' data analyses, it is likely these more accurate data analyses will result in better student-focused decisions, and thus help students.

Research Questions Q2b, Q3b, and Q4b. Research Questions Q2b, Q3b, and Q4b were all answered with insignificant findings concerning whether minor modifications in support format, mainly in terms of length and color usage, impacted educators' data analysis accuracy. In summary, these questions and their accepted hypotheses are featured below with question-specific implications, followed by implications that relate to all three research questions.

Q2b. Research Question Q2b was asked as follows:

- What impact does the manner in which a footer is framed, in terms of moderate differences in length and text color, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?

The following null hypothesis ($H2b_0$) was accepted and the alternative hypothesis was rejected ($H2b_a$) for Q2b based on the significant findings reported in *Chapter 4: Findings*:

- The null hypothesis was that the manner in which a footer was framed, in terms of moderate differences in length and text color, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.

This is different than saying the manner in which a footer was framed did not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data. Rather, since it is already accepted the format of such tools *does* matter, generally-similar yet slightly-dissimilar footer formats were investigated in this study. See *Chapter 3: Research Method: Delimitations* for more details.

Participants receiving Footer A indicated they used the footers 75% of the time, whereas participants receiving Footer B indicated they used the footers 70% of the time. When Footer A participants indicated they did not use the available footers, their data analysis accuracy was 27%, whereas when Footer B participants indicated they did not use the available footers, their data analysis accuracy was 6%. All 30 Footer A participants, regardless of footer use, averaged a data analysis accuracy of 36%, whereas all 30 Footer B participants, regardless of footer use, averaged a data analysis accuracy of 32%, In cases where respondents indicated they used the available footer, data analysis accuracy was 33% for Footer A participants and 40% for Footer B participants.

These insignificant findings imply either of the two footer formats investigated in this study works equally well in improving educators' data analysis accuracy, and educators were equally likely to use either of the two footer formats investigated. Since the two formats generally varied in size and density, they offered a window of text quantity that can be used as a reference for real world implementation. For example, Footer A was shorter and slightly less wordy (1st report footer: 39 words, 186 characters without spaces, 224 characters with spaces; 2nd report footer: 34 words, 156 characters without spaces, 228 characters with spaces) than the alternatively-framed footers and contained headings that utilized text color with meaning. Footer B was longer and slightly wordier (1st report footer: 58 words, 269 characters without spaces, 324 characters with spaces; 2nd report footer: 42 words, 199 characters without spaces, 237 characters with spaces) than the alternatively-framed footers and contained no headings or colored text. Thus study findings imply effective footers can range from 34 to 58 words, 156 to 269 characters without spaces, and 224-324 characters with spaces, and can

adhere to either form of color usage or level of color usage somewhere between the two examples (see *Appendix C*).

The implementation guideline provided above is not to be mistaken as an assertion this is the only effective way to provide a footer, as it is likely not. Other footer formats not investigated in this study might also be effective. However, research measuring their specific impact on educators' data analysis accuracy would have to be performed in order to make such a conclusion. See *Chapter 5: Implications, Recommendations, and Conclusions: Recommendations: Education Research Community* for related research recommendations.

Q3b. Research Question Q3b was asked as follows:

- What impact does the manner in which an abstract is framed, in terms of moderate differences in density and header color, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?

The following null hypothesis ($H3b_0$) was accepted and the alternative hypothesis was rejected ($H3b_a$) for Q3b based on the significant findings reported in *Chapter 4: Findings*:

- The null hypothesis was that the manner in which an abstract was framed, in terms of moderate differences in density and header color, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.

This is different than saying the manner in which an abstract was framed did not have an impact on the frequency with which educators drew accurate conclusions concerning

student achievement data. Rather, since it is already accepted the format of such tools *does* matter, generally-similar yet slightly-dissimilar abstract formats were investigated in this study. See *Chapter 3: Research Method: Delimitations* for more details.

Participants receiving Abstract A indicated they used the abstracts 53% of the time, whereas participants receiving Abstract B indicated they used the abstracts 47% of the time. When Abstract A participants indicated they did not use the available abstracts, their data analysis accuracy was 11%, whereas when Abstract B participants indicated they did not use the available abstracts, their data analysis accuracy was 9%. All 30 Abstract A participants, regardless of abstract use, averaged a data analysis accuracy of 21%, whereas all 30 Abstract B participants, regardless of abstract use, averaged a data analysis accuracy of 24%, In cases where respondents indicated they used the available abstract, data analysis accuracy was 31% for Abstract A participants and 36% for Abstract B participants.

These insignificant findings imply either of the two abstract formats investigated in this study works equally well in improving educators' data analysis accuracy, and educators were equally likely to use either of the two abstract formats investigated. Since the two formats generally varied in density, they offered a window of text quantity that can be used as a reference for real world implementation. For example, Abstract A was less dense and contained less information than the alternatively-framed abstracts and utilized heading color with meaning. Abstract B was denser and contained more information than the alternatively-framed abstracts and did not utilize heading color with meaning. Thus study findings imply effective abstracts can range in density and color usage as somewhere between the two examples (see *Appendix C*).

The implementation guideline provided above is not to be mistaken as an assertion this is the only effective way to provide an abstract, as it is likely not. Other abstract formats not investigated in this study might also be effective. However, research measuring their specific impact on educators' data analysis accuracy would have to be performed in order to make such a conclusion. See *Chapter 5: Implications, Recommendations, and Conclusions: Recommendations: Education Research Community* for related research recommendations.

Q4b. Research Question Q4b was asked as follows:

- What impact does the manner in which an interpretation guide is framed, in terms of moderate differences in length and information quantity, have on its ability to impact the frequency with which educators draw accurate conclusions concerning student achievement data?

The following null hypothesis ($H4b_0$) was accepted and the alternative hypothesis was rejected ($H4b_a$) for Q4b based on the significant findings reported in *Chapter 4: Findings*:

- The null hypothesis was that the manner in which an interpretation guide was framed, in terms of moderate differences in length and information quantity, would not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data.

This is different than saying the manner in which an interpretation guide was framed did not have an impact on the frequency with which educators drew accurate conclusions concerning student achievement data. Rather, since it is already accepted the format of such tools *does* matter, generally-similar yet slightly-dissimilar interpretation guide

formats were investigated in this study. See *Chapter 3: Research Method: Delimitations* for more details.

Participants receiving Interpretation Guide A indicated they used the interpretation guides 52% of the time, and participants receiving Interpretation Guide B also indicated they used the interpretation guides 52% of the time. When Interpretation guide A participants indicated they did not use the available interpretation guides, their data analysis accuracy was 0%, whereas when Interpretation Guide B participants indicated they did not use the available interpretation guides, their data analysis accuracy was 3%. All 30 Interpretation Guide A participants, regardless of interpretation guide use, averaged a data analysis accuracy of 32%, whereas all 30 Interpretation Guide B participants, regardless of interpretation guide use, averaged a data analysis accuracy of 28%, In cases where respondents indicated they used the available interpretation guide, data analysis accuracy was 48% for Interpretation Guide A participants and also 48% for Interpretation Guide B participants.

These insignificant findings imply either of the two interpretation guide formats investigated in this study works equally well in improving educators' data analysis accuracy, and educators were equally likely to use either of the two interpretation guide formats investigated. Since the two formats generally varied in size and density, they offered a window of text quantity that can be used as a reference for real world implementation. For example, Interpretation Guide A was shorter and contained less information (two pages) than the alternatively-framed interpretation guides and utilized heading color with meaning. Interpretation Guide B was longer and slightly wordier (three pages) than the alternatively-framed interpretation guides and did not utilize

heading color with meaning. Thus study findings imply effective interpretation guides can range from two to three pages, with a similar level of density, and can adhere to either form of color usage or level of color usage somewhere between the two examples (see *Appendix C*).

The implementation guideline provided above is not to be mistaken as an assertion this is the only effective way to provide an interpretation guide, as it is likely not. Other interpretation guide formats not investigated in this study might also be effective. However, research measuring their specific impact on educators' data analysis accuracy would have to be performed in order to make such a conclusion. See *Chapter 5: Implications, Recommendations, and Conclusions: Recommendations: Education Research Community* for related research recommendations.

Research Questions Q5a, Q5b, Q5c, Q5d, Q5e, and Q5f. Research Questions Q5a, Q5b, Q5c, Q5d, Q5e, and Q5f served the sole role of informing implications addressed by the primary research questions, specifically Q1, Q2a, Q3a, and Q4a. Research Questions Q5a, Q5b, Q5c, Q5d, Q5e, and Q5f were asked to determine if educators' school site demographics played a significant role in educators' data analysis accuracy, as this could impact the success of the three data analysis supports investigated with the primary research questions. It was found that none of the supports investigated with secondary Research Questions Q5a, Q5b, Q5c, Q5d, Q5e, and Q5f had a significant impact on educators' data analysis accuracy. In summary, these secondary research questions and their accepted hypotheses are featured below with question-specific implications.

Q5a. Research Question Q5a was asked as follows:

- What impact does an educator's school site level type (i.e., elementary or secondary) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H5a_0$) was rejected and the following alternative hypothesis was accepted ($H5a_a$) for Q5a based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's school site level type (i.e., elementary or secondary) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.04 features results for all 211 study participants, disaggregated by school level type, and *Chapter 4: Findings* features further explanation. An educator's school site level type (i.e., elementary or secondary) does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. In addition, school level type does not have a significant impact on whether or not an educator uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study's findings, to improve educators' data analysis accuracy, are not significantly impacted by educators' school level type. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the school level types where the system and reports will be used.

Q5b. Research Question Q5b was asked as follows:

- What impact does an educator's school site level (i.e., elementary, middle/junior high, or high school) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H5b_0$) was rejected and the following alternative hypothesis was accepted ($H5b_a$) for Q5b based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's school site level (i.e., elementary, middle/junior high, or high school) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.05 features results for all 211 study participants, disaggregated by school level, and *Chapter 4: Findings* features further explanation. An educator's school site level (i.e., elementary, middle/junior high, or high school) does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, school level has some impact on whether or not an educator uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study's findings, to improve educators' data analysis accuracy, are not significantly impacted by educators' school level. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the school levels where the system and reports will be used. However, district implementation could be aided by added encouragement to use available supports.

Q5c. Research Question Q5c was asked as follows:

- What impact does an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H5c_0$) was rejected and the following alternative hypothesis was accepted ($H5c_a$) for Q5c based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's school site academic performance, as measured by the 2012 Growth Academic Performance Index (API), which is the California state accountability measure, would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.06 features results for all 211 study participants, disaggregated by academic achievement of students at school sites, and *Chapter 4: Findings* features further explanation. An educator's school site academic performance, as measured by the 2012 Growth API, which is the California state accountability measure, does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, API has some impact on whether or not an educator uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study's findings, to improve educators' data analysis accuracy, are not significantly impacted by educators' school API. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the school API Growth scores where the system and reports will be used. However, district implementation could be aided by added encouragement to use available supports.

Q5d. Research Question Q5d was asked as follows:

- What impact does an educator's school site English Learner (EL) population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H5d_0$) was rejected and the following alternative hypothesis was accepted ($H5d_a$) for Q5d based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's school site English Learner (EL) population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.07 features results for all 211 study participants, disaggregated by percent of the school site's students who are classified as English Learner (EL), sometimes also called English Language Learner (ELL), and *Chapter 4: Findings* features further explanation. An educator's school site EL population does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, EL population has some impact on whether or not an educator uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study's findings, to improve educators' data analysis accuracy, are not significantly impacted by educators' school EL population. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the EL population where the system and reports will be used. However, district implementation could be aided by added encouragement to use available supports.

Q5e. Research Question Q5e was asked as follows:

- What impact does an educator’s school site Socioeconomically Disadvantaged population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H5e_0$) was rejected and the following alternative hypothesis was accepted ($H5e_a$) for Q5e based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator’s school site Socioeconomically Disadvantaged population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.08 features results for all 211 study participants, disaggregated by percent of the school site’s students who are classified as Socioeconomically Disadvantaged, and *Chapter 4: Findings* features further explanation. An educator’s school site Socioeconomically Disadvantaged population does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, Socioeconomically Disadvantaged population has some impact on whether or not an educator uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study’s findings, to improve educators’ data analysis accuracy, are not significantly impacted by educators’ school Socioeconomically Disadvantaged population. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the Socioeconomically Disadvantaged population where the system and reports will be used. However, district implementation could be aided by added encouragement to use available supports.

Q5f. Research Question Q5f was asked as follows:

- What impact does an educators' school site Students with Disabilities population have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H5f_0$) was rejected and the following alternative hypothesis was accepted ($H5f_a$) for Q5f based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's school site Students with Disabilities population would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.09 features results for all 211 study participants, disaggregated by percent of the school site's students who are classified as Students with Disabilities, and *Chapter 4: Findings* features further explanation. An educator's school site Students with Disabilities population does not have a significant impact on the frequency of accurate conclusions he or she draws concerning student achievement data. However, Students with Disabilities population has some impact on whether or not an educator uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study's findings, to improve educators' data analysis accuracy, are not significantly impacted by educators' school Students with Disabilities population. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the Students with Disabilities population where the system and reports will be used. However, district implementation could be aided by added encouragement to use available supports.

Research Questions Q6a, Q6b, Q6c, Q6d, and Q6e. Research Questions Q6a, Q6b, Q6c, Q6d, and Q6e served the sole role of informing implications addressed by the primary research questions, specifically Q1, Q2a, Q3a, and Q4a. Research Questions Q6a, Q6b, Q6c, Q6d, and Q6e were asked to determine if educators' demographics played a significant role in educators' data analysis accuracy, as this could impact the success of the three data analysis supports investigated with the primary research questions. It was found that none of the supports investigated with secondary Research Questions Q6a, Q6b, Q6c, Q6d, and Q6e had a significant impact on educators' data analysis accuracy. In summary, these secondary research questions and their accepted hypotheses are featured below with question-specific implications.

Q6a. Research Question 6a was asked as follows:

- What impact does an educator's veteran status have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H6a_0$) was rejected and the following alternative hypothesis was accepted ($H6a_a$) for Q6a based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's veteran status would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.10 features results for all 211 study participants, disaggregated by veteran status in the form of how many years the participant had spent working as an educator, , and *Chapter 4: Findings* features further explanation. An educator's veteran status does not have a significant impact on the frequency of accurate conclusions he or she draws

concerning student achievement data. In addition, veteran status has no significant impact on whether or not an educator uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study's findings, to improve educators' data analysis accuracy, are not significantly impacted by educator's veteran status. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the veteran status of educators who will be using the data system and reports.

Q6b. Research Question 6b was asked as follows:

- What impact does an educator's current professional role (e.g., teacher, site/school administrator, etc.) have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H6b_0$) was rejected and the following alternative hypothesis was accepted ($H6b_a$) for Q6b based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's current professional role (e.g., teacher, site/school administrator, etc.) would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.11 features results for all 211 study participants, disaggregated by the educator's current professional role, and *Chapter 4: Findings* features further explanation. An educator's current professional role (e.g., teacher, site/school administrator, etc.) does not have an impact on the frequency of accurate conclusions he or she draws concerning student achievement data. In addition, role has no significant impact on whether or not an

educator uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study's findings, to improve educators' data analysis accuracy, are not significantly impacted by educator's role. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the role of educators who will be using the data system and reports.

Q6c. Research Question 6c was asked as follows:

- What impact does an educator's perception of his or her own data analysis proficiency impact the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H6c_0$) was rejected and the following alternative hypothesis was accepted ($H6c_a$) for Q6c based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's perception of his or her own data analysis proficiency would not be related to the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.12 features results for all 211 study participants, disaggregated by perception of data analysis proficiency in the form of how participants rated their proficiency at analyzing student performance data, and *Chapter 4: Findings* features further explanation. An educator's perception of his or her own data analysis proficiency is not related to the frequency of accurate conclusions he or she draws concerning student achievement data. In addition, perceived data analysis proficiency has no significant impact on whether or not an educator uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study's findings, to

improve educators' data analysis accuracy, are not significantly impacted by educator's perception of data analysis proficiency. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the perceptions of educators who will be using the data system and reports.

Q6d. Research Question 6d was asked as follows:

- What impact does an educator's professional development over the past year, devoted specifically to *how* to analyze student data, have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H6d_0$) was rejected and the following alternative hypothesis was accepted ($H6d_a$) for Q6d based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's professional development over the past year, devoted specifically to how to analyze student data, would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.13 features results for all 211 study participants, disaggregated by data analysis professional development in the form of how many hours of PD the participant had taken part in within the past 12 months that specifically focused on learning how to correctly interpret student data, and *Chapter 4: Findings* features further explanation. An educator's professional development over the past year, devoted specifically to how to analyze student data, does not have an impact on the frequency of accurate conclusions he or she draws concerning student achievement data. In addition, PD has no significant impact on whether or not an educator uses an analysis support. These findings imply the

benefits of the three data analysis supports shown, through this study's findings, to improve educators' data analysis accuracy, are not significantly impacted by educator's data analysis PD. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the data analysis PD of educators who will be using the data system and reports. This is not to be mistaken as an assertion that data analysis PD is not needed or beneficial, as neither assertion would be accurate.

Q6e. Research Question 6e was asked as follows:

- What impact does the number of graduate-level educational measurement courses an educator has taken have on the frequency with which he or she draws accurate conclusions concerning student achievement data?

The null hypothesis ($H6e_0$) was rejected and the following alternative hypothesis was accepted ($H6e_a$) for Q6e based on the significant findings reported in *Chapter 4: Findings*:

- The alternative hypothesis was that an educator's number of graduate-level educational measurement courses would not have an impact on the frequency of accurate conclusions he or she drew concerning student achievement data.

Table 4.14 features results for all 211 study participants, disaggregated by educational measurement course number in the form of how many graduate-level courses the participant had taken that were specifically dedicated to educational measurement, and *Chapter 4: Findings* features further explanation. An educator's number of graduate-level educational measurement courses does not have an impact on the frequency of accurate conclusions he or she draws concerning student achievement data. In addition, graduate educational measurement courses has no significant impact on whether or not an educator

uses an analysis support. These findings imply the benefits of the three data analysis supports shown, through this study's findings, to improve educators' data analysis accuracy, are not significantly impacted by educator's graduate educational measurement courses. Thus those implementing these supports in their data systems and reports can expect similar success regardless of the graduate educational measurement education of educators who will be using the data system and reports. This is not to be mistaken as an assertion that graduate educational measurement courses are not needed or beneficial, as neither assertion would be accurate.

Limitations. The study dealt exclusively with educators and their use of data system reports and resources in an isolated setting. Thus, to maintain external validity, study findings may not be applied to inferences concerning non-educators, such as parents, students, or politicians. Likewise, in consideration of the potential impact of interaction of setting and treatment, no generalizations of data analyses may be made of analysis environments that are not report-based, such as data analyses made based on data group discussions or based on an explanation heard by a data coach.

Contributions to existing literature in the field. The FDA directs the pharmaceutical industry to accompany over-the-counter medication with textual guidance regarding its use but to also provide solid evidence on how effective its labeling is in reducing errors; to proceed without such research is considered negligent (DeWalt, 2010). Despite the common use of data systems to generate reports, research on aspects of report format and system support that could enhance analysis accuracy is scarce (Goodman & Hambleton, 2004). As covered in *Chapter 2: Literature* Review, Research that was devoted to data system and report format, including how effectively this format

communicates data to users, previously focused on participants' preferences and participants' perceived value of supports. However, user preference can be the opposite of the reporting format that actually renders the more accurate interpretation (Hattie, 2010).

This study was used to measure specifically how effective varied analysis supports are in improving data analysis accuracy, and it did not rely on participants' preferences or perceived value of supports. The findings of this study fill a void in education field literature by containing evidence that can be used to identify:

- whether data systems can help increase data analysis accuracy by providing analysis support within data systems and their reports, with the finding being that they can.
- three specific data system/report-embedded supports that increase educators' data analysis accuracy.
- the specific degree to which these supports increase educators' data analysis accuracy (*Figure 4.01*), with results disaggregated by educator and site demographics and by reporting environment.
- how likely educators are to use each support, disaggregated by educator and site demographics and by reporting environment.
- examples showing what effective footers, abstracts, and interpretation guides look like (*Appendix C*).
- whether minor modifications in support format, mainly in terms of length and color usage, impacted educators' data analysis accuracy, with the findings being that differences in data analysis accuracy were insignificant.

Improvements data system and report providers make in light of this study have the potential to improve the accuracy with which educators analyze the data generated by their data systems. More accurate data analyses will likely result in more accurate data-informed decision-making for the benefit of students.

Recommendations

This study warranted recommendations for three key roles: (a) data system and report providers, such as data system vendors and also district staff who maintain in-house data systems; (b) educators who use data systems and reports, particularly those in leadership positions who play a role in data system selection, support, and replacement; and (c) the education research community. These recommendations, which include recommendations for future research, are based on the following key findings, all of which were found to be significant:

- Educators' data analyses were 264% more accurate (with an 18 percentage point difference) when any one of the three supports – footer, abstract, or interpretation guide – was present, and 355% more accurate (with a 28 percentage point difference) when respondents specifically indicated having used the support.
- Educators' data analyses were 307% more accurate (with a 23 percentage point difference) when a footer was present, and 336% more accurate (with a 26 percentage point difference) when respondents specifically indicated having used the footer.
- Educators' data analyses were 205% more accurate (with a 12 percentage point difference) when an abstract was present, and 300% more accurate (with a 22

percentage point difference) when respondents specifically indicated having used the abstract.

- Educators' data analyses were 273% more accurate (with a 19 percentage point difference) when an interpretation guide was present, and 436% more accurate (with a 37 percentage point difference) when respondents specifically indicated having used the interpretation guide.

The specifics of the above findings are detailed in *Chapter 4: Findings*. As explained earlier in *Chapter 5: Implications, Recommendations, and Conclusions: Implications*, it has now been proven, through this study, that accompanying data reports with footers, abstracts, and interpretation guides is beneficial to educators' data analyses and thus beneficial to the students affected by educators data-informed decisions.

Data system and report providers. Data system and report providers, such as data system vendors and also district staff who maintain in-house data systems, are encouraged to create a report-specific footer, abstract, and interpretation guide for each of the reports they provide. At the very least, they should provide one of these three supports. The footer would be a good starting point, as it is most likely the fastest and easiest to implement while also being highly effective.

Support contents. Each support was designed, through its contents and the clarity of its delivery of that contents, to prevent wrong analyses, such as preventing common analysis mistakes specific to the report's particular datasets, while providing guidance on the data's accurate analyses. Thus it is crucial that the person or people who determine the contents of each support:

- are well-versed in region-specific data and its analyses,

- have led many educators of varied roles and backgrounds in the analyses of this data and are thus well-versed in the most common mistakes made when analyzing it,
- have an educator background in order to know how to best communicate with educators of varied roles and backgrounds, such as knowing terms, approaches, and verbiages to use versus not use.

Support format. When this content is added to the final report or support, the person assembling the final product should have a thorough understanding of how design impacts understanding. Reading this complete dissertation can help to provide such an understanding, but a stronger design education and background is recommended. For example, someone building an abstract should know how to use white space rather than cramming all of the text into the top half of the page. The handouts used in this study (*Appendix C*) can be used as references, as can their specifications based on findings explained in the Q2b, Q3b, and Q4b sections of *Chapter 5: Implications, Recommendations, and Conclusions: Implications*:

- Footers should be located at the bottom of reports in the same font size as the majority of the report's data; can range from 34 to 58 words, 156 to 269 characters without spaces, and 224-324 characters with spaces; and can adhere to either form of color usage or level of color usage somewhere between the two examples (see *Appendix C*).
- Abstracts should contain similar contents and range in density and color usage as somewhere between the two examples (see *Appendix C*).

- Interpretation guides should range from two to three pages, with a similar level of density, and should contain similar contents while adhering to either form of color usage or level of color usage somewhere between the two examples (see *Appendix C*).

The implementation guidelines provided above are likely not the only effective ways to provide supports, as other formats not investigated in this study might also be effective. However, research measuring their specific impact on educators' data analysis accuracy would have to be performed in order to make such conclusions. See *Chapter 5: Implications, Recommendations, and Conclusions: Recommendations: Education Research Community* for related research recommendations.

For people's convenience and in order to promote the effective formats established with this study, the researcher has created templates for abstracts and interpretation guides and has housed them online to be accessed by anyone wanting to use them. Note these templates are provided in both docx and doc formats in order to accommodate both PC an d Mac users, as well as users of both older and newer software:

- www.overthecounterdata.com/s/AbstractTemplates.docx for PC users of Microsoft® Office 2007 or later (or else the files will not display correctly) and www.overthecounterdata.com/s/AbstractTemplates.doc for Mac users or PC users of older versions of Microsoft® Office; each file contains two templates: one for a simpler abstract version and one for a denser abstract version
- www.overthecounterdata.com/s/IntGuideTemplates.docx for PC users of Microsoft® Office 2007 or later (or else the files will not display correctly) and www.overthecounterdata.com/s/IntGuideTemplates.doc for Mac users or PC

users of older versions of Microsoft® Office; each file contains two templates: one for a shorter interpretation guide version and one for a longer interpretation guide version

See Appendix R for an image of each of the four templates provided.

Support access. The supports investigated in this study were found to be effective when provided to educators in concert with the report they explained. Footers are provided directly at the bottom of reports and are thus provided to educators in tandem with the reports. However, abstracts and interpretation guides constitute separate, supplemental documentation. Thus the data system and report provider must take steps to ensure an educator viewing a report is simultaneously provided with access to the report's abstract and interpretation guide.

While some educators (44%) use the data system directly, others (56%) have access but do not use the data system directly and instead only read printed versions of reports others used the data system to generate (Underwood et al., 2008). Thus the data system and report provider must provide the supplemental documentation in ways that account for both types of users: online versus printed. Examples of how these two user types can be accommodated include:

- two links visible and accessible while viewing the report online, within the data system: an "Abstract" link leading directly to the report's abstract and an "Interpretation Guide" link leading directly to the report's interpretation guide
- the same solution explained above, with the added stipulation that the abstract and interpretation guide each be downloadable as an Adobe pdf file so users can print it to provide it to others when reports are provided in printed form, attach it to an

email to send it to others when reports are provided in emailed form, save it somewhere such as a staff portal for others where already-generated reports are housed, or other approach to sharing its access

If the data system has a help system, this is another location where links to abstracts and interpretation guides can appear to increase awareness of them and access to them.

Educators. Educators who use data systems and reports, particularly those in leadership positions who play a role in data system selection, support, and replacement, are in positions to encourage their data system and report providers to add footers, abstracts, and interpretation guides to accompany reports they offer. There are five key reasons educators should request such supports and promote awareness and dialogue about such supports in educator communities:

- Each of the three supports results in a significant increase in educators' data analysis accuracy when the supports are present, as noted above.
- Without supports, educators' data analysis accuracy is only 11% correct.
- Educators use their data analyses to inform decisions that impact students, so providing such supports is of dire import to students' wellbeing.
- Unlike popular approaches currently used by educators to improve their own data analysis accuracy, such as PD and staff supports, adding footers, abstracts, and interpretation guides to data systems will likely not cost educators any money, as it is recommended that data system and report providers incur any related costs. One likely exception to this rule is a district where an in-house data system is maintained.

- Popular approaches currently used by educators to improve their own data analysis accuracy, such as PD and staff supports, are not omnipotent. While evidence indicates these approaches are beneficial, they do not result in fool-proof data analyses. Rather, each approach has some limitations. See *Chapter 2: Literature Review: Supports Outside of Data Systems Are Not Enough* for details. Thus added supports, such as those examined with this study, are needed.

It is thus recommended educators take the following steps, based on which steps are appropriate for their roles and circumstances, to capitalize on the benefits of the three over-the-counter data supports investigated in this study:

- Encourage current data system and report providers to add footers, abstracts, and interpretation guides to accompany reports they offer. This dissertation and its findings can be used as support for requests.
- Add footers, abstracts, and interpretation guides as consideration criteria when deciding whether to keep or purchase/hire a data system or report provider. For example, when issuing a request for proposal (RFP) inviting vendors to submit proposals for their data systems, add these supports as required or desired criteria and consider their presence in the selection process. This could encourage vendors to add these supports and/or could raise their awareness of the need to add such supports.
- Promote awareness and dialogue about such supports in educator communities. For example, discuss the importance of footers, abstracts, and interpretation guides with colleagues and share approaches to acquiring and using them.

- If a report was created outside of a data system, or if the data system does not contain over-the-counter data supports, personally add a footer to the report and create an abstract and interpretation guide to go with it following guidance provided earlier in this chapter.

Education research community. There are two main topics recommended for future research on the topic of over-the-counter data supports such as footers, abstracts, and interpretation guides. Education research community members are encouraged to fill remaining gaps in field literature by investigating the topics that follow. Each of these should be studied in terms of specific impact on educators' data analysis accuracy as opposed to which supports and formats educators prefer. As noted, user preference can be the opposite of the reporting format that actually renders the more accurate interpretation (Hattie, 2010).

All supports simultaneously. Although the three supports investigated in this study increased educators' data analysis accuracy by 205%-307% when they were merely present, and by even more – 300%-436% – when participants specifically indicated they used them (see *Figure 4.01*), no single support resulted in 100% data analysis accuracy for *all* users. The average data analysis accuracy only rose to up to 48% correct (see *Table 4.02*). This was expected, mainly due to a key assumption made in this study.

Assumptions about the study population included that respondents would make reasonable attempts to answer the four data analysis questions – Questions 4-7 – correctly, but they would not necessarily answer the questions to the best of their abilities. Because most survey completion sessions were conducted at the end of the school day, which meant at the end of each participant's work day, it was reasonable to

assume respondents were tired, which is not conducive to data analysis accuracy. For example, fatigue at the end of a workday can cause a significant decline in interpretation accuracy (Krupinski & Berbaum, 2010). However, the times when these survey sessions were conducted – when staff members were not teaching – were also the time these educators would be most likely to have the time to conduct their real-life analyses of student data.

Since it is important to continually look for ways educators' data analysis accuracy can be improved even more, particularly in ways that are not intrusive on educator time or resources, one can imagine the likely-beneficial impact of offering educators all three supports as opposed to just one. 87% of control group participants who had no access to supports indicated they would have used the added support if they had it, and 58% of participants who had supports indicated they used them. As explained earlier, there is reason to suspect this 58% was even higher due to the significant impact of support presence even when respondents indicated they did not use them. Even if the 58% is accurate, it means some educators opted not to use the supports. However, one can imagine that different supports – which range in size, format, quantity of information, and more – appeal to different educators. 73% of participants with footers used them, 50% of participants with abstracts used them, and 52% of participants with interpretation guides used them (see *Table 4.02*).

One recommendation for further study is thus to measure the impact on educators' data analyses when the support of footers, abstracts, and interpretation guides are offered to educators in concert. Likewise, the impact of all over-the-counter data components used jointly is worthy of further research: label, supplemental documentation, help

system, package/display, and contents. For example, the package/display of reports used in the study did not fully reflect likely-ideal reporting practices, as the study's reports needed to closely reflect report formats in current use by typical data systems, which are not always ideal; thus the format used for the study's Report 1 is highly typical despite the fact that it would possibly render better analyses if it was displayed as shown in *Figure 5.01*. While a lab-environment is recommended, where variables and conditions can be controlled, data from real-world, non-lab environments would also be valuable in cases where these supports are implemented in data systems used by one or more school districts.

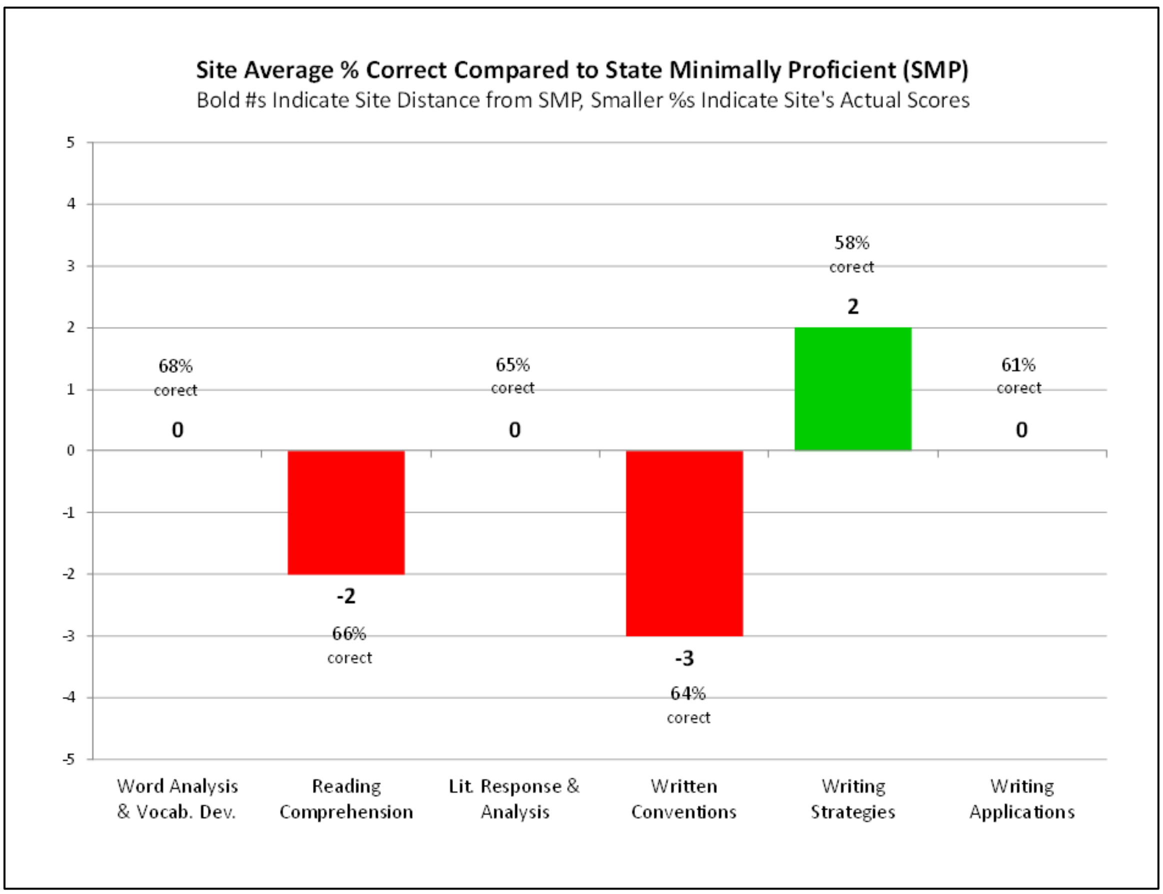


Figure 23: *Likely More Effective Format for Report 1 yet Atypical of Data Systems*

Additional formats. The minor modifications in support format investigated in this study, mainly in terms of length and color usage, had no significant impact on the participating educators' data analysis accuracy. This resulted in acceptance of the null hypotheses for primary Research Questions Q2b, Q3b, and Q4b. These results were somewhat unexpected given literature on behavioral economics, particularly in the area of framing, and literature on report and documentation design. However, it is important to note all support format variations used in the study subscribed to best practices recommended in literature on report and documentation design. Thus the variations were minor and designed to garner more specificity in these best practices. It was thus concluded such minor variations are equally minor in their impact on educators' data analyses.

Nonetheless, the manner in which content is organized for people using it to make decisions significantly impacts those decisions (Thaler & Sunstein, 2008). Framing applies to how information is presented, as presenting the same information to someone in different ways will often result in different emotions and different levels of difficulty in understanding or analyzing the information (Kahneman, 2003, 2011). Thus suggested ways to present analysis guidance in footers, abstracts, and interpretation guides were utilized in the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study, but the best manner in which to frame these resources had not yet been determined in regards to direct impact on analysis accuracy. Thus each of the three support resources were framed in two different formats for respondents in the study. Both formats, in each case, were found to be equally effective. However, it would be premature and possibly

wrong to conclude the two formats used in this study, while significantly effective, were the *most* effective formats possible.

It is recommended the education research community continue to explore additional formats for footers, abstracts, and interpretation guides in order to continually search for better ways to provide added analysis support to educators. Likewise, other over-the-counter data aspects such as non-footer aspects of report labeling, the data system's help system, report packaging and data display, and report contents should also be investigated in order to inform better data systems and reports that provide optimal support with educators' data analyses.

Conclusions

Most educators have access to data systems to generate and analyze score reports (Aarons, 2009; Herbert, 2011). However, educators do not use this data correctly, and there is clear evidence many users of data system reports have trouble understanding the data (Wayman et al., 2010; Zwick et al., 2008). The impact of feedback, which is considered one of the most powerful influences on student learning and achievement, can be negative if the performance feedback is not provided in the best way (Hattie & Timperley, 2013). Despite this, labeling and tools within data systems to assist analysis are uncommon (USDEOPEPD, 2009). The *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study rendered findings that data system-embedded data analysis support in the forms of footers, abstracts, and interpretation guides all have a significant, positive impact on the accuracy of educators' data analyses.

Findings rendered implications there are direct benefits to educators' data use when a data system and its reports embed at least one of the three data analysis supports

investigated in this study. Findings also supported experts' assertions that educators desire more data analysis support from their data systems and its reports, and that the majority of educators use such supports when they are available. In addition, secondary research questions concerning educators' personal and school site demographics were answered with the finding that such demographics have no significant bearing on the supports' success, and thus the supports can be implemented with expected success at varied locations and for varied users.

Given the significant success of footers, abstracts, and interpretation guides, the study warranted related recommendations for three key roles: (a) data system and report providers, such as data system vendors and also district staff who maintain in-house data systems; (b) educators who use data systems and reports, particularly those in leadership positions who play a role in data system selection, support, and replacement; and (c) the education research community. The last of these was paired with recommendations for future research, mainly in terms of (a) testing the success of all three supports when provided in concert, and (b) investigating additional formats for footers, abstracts, and interpretation guides in order to continually search for better ways to provide added analysis support to educators. Likewise, the education research community was encouraged to explore best practices for other over-the-counter data aspects such as non-footer aspects of report labeling, the data system's help system, report packaging and data display, and report contents in order to inform better data systems and reports that provide optimal support with educators' data analyses.

The findings of the *Over-the-Counter Data's Impact on Educators' Data Analysis Accuracy* study fill a void in education field literature by containing evidence that can be used to identify:

- whether data systems can help increase data analysis accuracy by providing analysis support within data systems and their reports, with the finding being that they can.
- three specific data system/report-embedded supports that increase educators' data analysis accuracy.
- the specific degree to which these supports increase educators' data analysis accuracy (*Figure 4.01*), with results disaggregated by educator and site demographics and by reporting environment.
- how likely educators are to use each support, disaggregated by educator and site demographics and by reporting environment.
- examples showing what effective footers, abstracts, and interpretation guides look like (*Appendix C*).
- whether minor modifications in support format, mainly in terms of length and color usage, impacted educators' data analysis accuracy, with the findings being that differences in data analysis accuracy were insignificant.

Improvements data system and report providers make in light of this study have the potential to improve the accuracy with which educators analyze the data generated by their data systems. More accurate data analyses will likely result in more accurate data-informed decision-making for the benefit of students. It is the strong conviction of this researcher that students deserve for stakeholders to use *all* possible supports for improved

analysis accuracy in an effort to completely eliminate – rather than merely reduce – their data analysis errors, as these errors impact decisions that impact students' lives.

References

Aarons, D. (2009). Report finds states on course to build pupil-data systems. *Education Week*, *29*(13), 6. Retrieved from http://search.proquest.com/docview/202710770?accountid=28180

Airasian, P. (2000). *Assessment in the classroom: A concise approach* (2nd ed.). Boston, MA: McGraw Hill.

Aladjem, D. K., LeFloch, K. C., Zhang, Y., Kurki, A., Boyle, A., Taylor, J. E., Hermann, S., Uekawa, K., Thomsen, K., & Fashola, O. (2006). *Models matter—The final report of the National Longitudinal Evaluation of Comprehensive School Reform*. Washington, D.C.: American Institutes for Research (ERIC Document Reproduction Service No. ED499198)

Allalouf, A. (2007). An NCME instructional module on quality control procedures in the scoring, equating, and reporting of test scores. *Educational Measurement: Issues and Practice, 26(1)*, 36-46.

Alverson, C. Y. (2008). *Exploring differences in teachers', administrators', and parents' preferences for data display and whether type of graphic display influences accuracy when extracting information*. University of Oregon. ProQuest Dissertations and Theses, Retrieved from http://search.proquest.com/docview/304502112?accountid=28180

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Institutes for Research (AIR). (2013). *Most teachers "highly qualified" under NCLB standards, but teacher qualifications lag in many high poverty and high minority schools*. Retrieved from http://www.air.org/reports-products/index.cfm?fa=viewContent&content_id=417.

American Psychological Association (2001). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Armstrong, J., & Anthes, K. (2001). How data can help. *American School Board Journal, 188(11)*, 38-41.

Aschbacher, P. R., & Herman, J. L. (1991). *Guidelines for effective score reporting: CSE Technical Report 326*. Los Angeles, CA: UCLA Center for Research on Evaluation, Standards, and Student Testing.

Assessment and Accountability Division of California Department of Education (2011). *2010–11 Title III accountability report information guide.* Retrieved from http://www.cde.ca.gov/ta/ac/t3/documents/infoguide1011.pdf

Benadom, E. (2005). How "smart data" helps English learners. *Leadership, 34*(3), 26. Retrieved from http://proquest.umi.com.proxy1.ncu.edu/pqdweb?did=782951011&Fmt=7&clientId=52110&RQT=309&VName=PQD and http://proquest.umi.com.proxy1.ncu.edu/pqdweb?vinst=PROD&fmt=6&startpage=-1&clientid=52110&vname=PQD&RQT=309&did=782951011&scaling=FULL&vtype=PQD&rqt=309&TS=1298830994&clientId=52110

Bennett, R. E., & Gitomer, D. H. (2009). Transforming K-12 assessment: Integrating accountability testing, formative assessment and professional support. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century*, 43-61. New York, NY: Springer.

Bernhardt, V. L. (2004). *Data analysis for continuous school improvement.* Larchmont, NY: Eye on Education.

Bernhardt, V. L. (2007). *Translating data into information to improve teaching and learning.* Larchmont, NY: Eye On Education.

Bill and Melinda Gates Foundation (2007). *Landscape review: Education data.* Retrieved from http://www.gatesfoundation.org/learning/Documents/landscape-review-education-data.pdf

Bill and Melinda Gates Foundation (2012). *Innovation in education: Technology and effective teaching in the U.S.* Retrieved from https://edsurge.s3.amazonaws.com/public/BMGF_Innovation_In_Education.pdf

Black, T. R. (1999). *Doing quantitative research in the social sciences: An integrated approach to research design, measurement, and statistics*. Thousand Oaks, CA: Sage Publications, Inc.

Brown-Brumfield, D., & DeLeon, A. (2010). Adherence to a medication safety protocol: Current practice for labeling medications and solutions on the sterile field. *Association of Operating Room Nurses*. AORN Journal, 91(5), 610-610-7. doi:10.1016/j.aorn.2010.03.002

California Department of Education. (2011). *California Standards Tests technical report.* Concord, CA: Educational Testing Service.

California Department of Education (2012a). *2011–12 Accountability Progress Reporting System: 2011–12 Title III Accountability Report information guide*. Retrieved from http://www.cde.ca.gov/ta/ac/t3/documents/infoguide11-12.pdf

California Department of Education. (2012b). *2012 California Standardized Testing and Reporting: Post-test guide technical information for STAR district and test site coordinators and research specialists*. Concord, CA: Educational Testing Service.

Camerer, C., Loewenstein, G., & Rabin, M. (2003). *Advances in behavioral economics*. Princeton, NJ: Princeton University Press.

Cho, V., & Wayman, J. C. (2009, April). *Knowledge management and educational data use.* Paper presented at the 2009 Annual Meeting of the American Educational Research Association, San Diego, CA.

Clay, B. (2012). House bills introduced to compel mandatory labeling for GMO foods. *The Canary Party*. Retrieved from http://canaryparty.net/index.php/the-news/81

Coburn, C. E., Honig, M. I., & Stein, M. K. (2009). What's the evidence on districts' use of evidence? In J. Bransford, D. J. Stipek, N. J. Vye, L. Gomez, & D. Lam (Eds.), *The role of research in educational improvement*, 67-88. Cambridge, MA: Harvard Education Press.

Coburn, C. E., Toure, J., & Yamashita, M. (2009). Evidence, interpretation, and persuasion: Instructional decision making in the district central office. *Teachers College Record*, 111, 1115–1161.

Combs, G. (2004). Why teachers hate tech training... and what to do about it. *MultiMedia & Internet@Schools, 11*(1), 8-8.

Data Quality Campaign (2009). *The next step: Using longitudinal data systems to improve student success*. Retrieved from http://www.dataqualitycampaign.org/find-resources/the-next-step/

Data Quality Campaign (2010). *Creating reports using longitudinal data: how states can present information to support student learning and school system improvement*. Retrieved from http://www.dataqualitycampaign.org/files/DQC%20Reports%20Issue%20Brief%20Nov8.pdf

Data Quality Campaign (2011). *Leveraging the power of state longitudinal data systems: Building capacity to turn data into useful information*. Retrieved from http://www.dataqualitycampaign.org/files/DQC-Research%20capacity%20May17.pdf

Day, B. (2010). *The effects of expectations on performance: Generalizing galatea*. The University of Tennessee at Chattanooga. ProQuest Dissertations and Theses, Retrieved from http://search.proquest.com/docview/851699033?accountid=28180

De Jong, J. H. A. L. & Zheng, Y. (2011). *Research Note: Applying EALTA Guidelines: A Practical case study on Pearson Test of English Academic*. London, UK: Pearson.

DeWalt, D. A. (2010). Ensuring safe and effective use of medication and health care: perfecting the dismount. *The Journal of the American Medical Association (JAMA), 304(23)*, 2641-2642. doi: 10.1001/jama.2010.1844

Fast, E. & State Collaborative on Assessment and Student Standards Accountability Systems and Reporting Consortium. (2002, December). *A guide to effective*

accountability reporting. Washington, DC: Council of Chief State School Officers.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Few, S. (2006, February). *Common pitfalls in dashboard design*. Boise, ID: ProClarity Corporation.

Few, S. (2008, November 14). Telling compelling stories with numbers: Data visualization for enlightening communication. *Statewide Longitudinal Data Systems (SLDS) Grant Program Third Annual Fall Grantee Meeting.* Presentation conducted from SLDS, Arlington, VA. Retrieved from http://nces.ed.gov/programs/slds/pdf/08_F_06.pdf

Fiore, R. N. (2011). *Conflicts of interest in research involving human subjects*. Retrieved from https://www.citiprogram.org/members/learnersII/moduletext.asp?strKeyID=4814E3F2-528A-4C6B-AE1F-86AF254DC861-8882253&module=488

Fletcher, G. H. (January, 2012). The water bed effect in k-12 education. *The Journal, 39*(1), 18-19.

Gall, M. D., Borg, W. R., & Gall, J. P. (2007). *Educational research: An introduction* (8th ed.). Boston, MA: Allyn & Bacon.

Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education, 17*(2), 145–220.

Gorman, S., & National Center for Education Statistics (2010). *An introduction to NAEP: National Assessment of Educational Progress*, NCES 2010468. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Gross, B., & Goertz, M. E. (Eds.). (2005). *Holding high hopes: How high schools respond to state accountability policies* (CPRE Research Report No. RR-056). Philadelphia, PA: Consortium for Policy Research in Education.

Halpin, J., & Cauthen, L. (July 31, 2011). The education dashboard. *Center for Digital Education's Converge Special Report 2*(3), 2-36.

Hambleton, R. K. (2002). How can we make NAEP and state test score reporting scales and reports more understandable? In R. W. Lissitz & W. D. Schafer (Eds.), *Assessment in educational reform*, 192-205. Boston, MA: Allyn & Bacon.

Hambleton, R. K., & Slater, S.C. (1996). *Are NAEP executive summary reports understandable to policymakers and educators?* Paper presented at the annual meeting of the National Council on Measurement in Education, New York. (ERIC Document Reproduction Service No. ED400296)

Hamilton, L. S., & Koretz, D. M. (2002). Tests and their use in test-based accountability systems. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.), *Making sense of test-based accountability in education*, 13-49. Santa Monica, CA: Rand.

Hampton, T. (2007). Groups urge warning label for medical devices containing toxic chemical. *The Journal of the American Medical Association (JAMA), 298(11)*, 1267. doi: 10.1001/jama.298.11.1267

Harris, R. L., (1999). *Information graphics: A comprehensive illustrated reference*. New York, NY: Oxford University Press.

Hattie, J. (2009, April). *Visibly learning from reports: The validity of score reports*. Paper presented at the annual meeting of the National Council for Measurement in Education (NCME), San Diego, CA.

Hattie, J. (2010). Visibly learning from reports: The validity of score reports. *Online Educational Research Journal*. Also: Paper presented at the annual meeting of the National Council for Measurement in Education (NCME), San Diego, CA. Retrieved from http://www.oerj.org/View?action=viewPaper&paper=6

Hattie, J. A. C., & Brown, G. T. L. (2008). Technology for school-based assessment and assessment for learning: Development principals from New Zealand. *Journal of Educational Technology Systems 36*(2), 189-201.

Hattie, J., & Hargadon, S. (2013, April 3). John Hattie on what actually works in schools to improve learning. *Thursday Interview with Steve Hargadon*. Presentation conducted from Blackboard Collaborate (formerly Elluminate). Attended at https://sas.elluminate.com/site/external/jwsdetect/playback.jnlp?psid=2013-04-04.1504.M.9E9FE58134BE68C3B413F24B3586CF.vcr&sid=2008350

Hattie, J., & Timperley, H. (2007, March). The power of feedback. *Review of Educational Research, 77*(1), 81-112. doi: 10.3102/003465430298487

Herbert, M. (2011). States on track to have top-notch data systems. *District Administration, 47(4),* 12. Norwalk, CT: Professional Media Group LLC.

Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (November, 2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning and Assessment, 5(2)*. Retrieved from http://escholarship.bc.edu/jtla/vol5/2/

Illuminate Education (2012). Prebuilt Report Abstracts. *Illuminate Help Manuals: EXTRAS: Report Abstracts (Guides)*. Retrieved from http://help-isi.illuminateed.com/s/dna_isi_help_manual/m/5571/l/52729-prebuilt-report-abstracts

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106(6)*, 1258–1287.

Inoue, H. & Tagawa, T. (2006). On the Construction of Intelligent Online-Help System for Authoring Course Materials on Course Management System. In C. Crawford et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2006*, 1628-1633. Chesapeake, VA: AACE.

Jaeger, R. (1992). General issues in reporting of the NAEP trial state assessment results. In R. Glaser & R. Linn (Eds.), *Assessing student achievement in the states*, 107-109. Stanford, CA: National Academy of Education.

Jaeger, R. M. (2003). *NAEP validity studies: reporting the results of the National Assessment of Educational Progress*, NCES 2003–11. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Johnson, C. A. (2012). *The information diet: A case for conscious consumption*. Sebastopol, CA: O'Reilly Media, Inc.

Johnson, E. J., Hershey, J., Meszaros, J., Kunreuther, H. (1993, August). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty, 7*(1), 35-51.

Joint Committee on Testing Practices (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association. Retrieved from onhttp://www.apa.org/science/programs/testing/fair-code.aspx

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review, 93(5)*, 1449-1475.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (Ed.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Kenny, D. (2010). DataToText: A consumer-oriented approach to data analysis. Multivariate Behavioral Research, 45: 6, 1000-1019. doi:10.1080/00273171.2010.531232

Killion, J., & Hirsh, S. (2012, February). The bottom line on excellence: A guide to investing in professional learning that increases educator performance and student results. *JSD: The Learning Forward Journal, 33*(1). Oxford, OH: Learning Forward.

Knapp, M. S., Swinnerton, J. A., Copland, M. A., & Monpas-Hubar, J. (2006). *Data-informed leadership in education*. Seattle, WA: Center for the Study of Teaching and Policy.

Knight Commission on the Information Needs of Communities in a Democracy (2009). *Informing communities: Sustaining democracy in the digital age*. Washington, D.C.: The Aspen Institute.

Koretz, D., & Deibert, E. (1993). *Interpretations of National Assessment of Educational Progress (NAEP) anchor points and achievement levels by the print media in 1991*. Santa Monica, CA: Rand.

Krupinski, E. A., & Berbaum, K. S. (2010, February 23). Does reader visual fatigue impact interpretation accuracy?. *Proc. SPIE 7627, Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment, 76270M*. doi:10.1117/12.841050.

Kuehn, B. M. (2009). FDA focuses on drugs and liver damage: Labeling and other changes for acetaminophen. *The Journal of the American Medical Association (JAMA), 302(4)*, 369-371. doi: 10.1001/jama.2009.1019

Lachat, M. A., & Smith, S. (2005, July). Practices that support data use in urban high schools. *Journal of Education for Students Placed At Risk, 10(3)*, 333-349.

Leeson, H. V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing, 6(1)*, 1-24

Lehrer, J. (2011). Is 'nudging' really enough? *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB10001424052748703385404576259030599041442.html

Lewis, D., Madison-Harris, R., Muoneke, A., & Times, C. (2010). Using data to guide instruction and improve student learning. *SEDL Letter, 22*(2), 10–12.

Light, D., Wexler, D. & Heinz, C. (2005). Keeping teachers in the center: a framework for data-driven decision-making. In C. Crawford et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2005*, 128-133. Chesapeake, VA: AACE.

Linn, R. L., & Dunbar, S. B. (1992). Issues in the design and reporting of the National Assessment of Educational Progress. *Journal of Educational Measurement, 29*(2), 177-194.

Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice, 23(2)*, 26–32.

Lyrén, P. (2009). Reporting subscores from college admission tests. *Practical Assessment, Research & Evaluation, 14(4)*, 3-12. Retrieved from http://pareonline.net/pdf/v14n4.pdf

Marlow, M. L., & Abdukadirov, S. (2012). Can behavioral economics combat obesity? *Regulation, 35*(2), 14-18. Retrieved from http://search.proquest.com/docview/1030094290?accountid=28180

Marsh, J. A., McCombs, J. S., & Martorell, F. (2010). How instructional coaches support data-driven decision making. *Educational Policy, 24(6)*, 872-907.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education: Evidence from recent RAND research*. Santa Monica, CA: RAND Corporation.

McDonald, S., Andal, J., Brown, K., & Schneider, B. (2007). *Getting the evidence for evidence-based initiatives: How the Midwest states use data systems to improve education processes and outcomes (Issues & Answers Report, REL 2007–No. 016)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Assistance, Regional Educational Laboratory Midwest.

McLaughlin, M., & Talbert, J. E. (2006). *Building school-based teacher learning communities: Professional strategies to improve student achievement*. New York, NY: Teachers College Press.

Metropolitan Life Insurance Company. (2013). *MetLife survey of the American teacher: Challenges for school leadership.* New York, NY: Author and Peanuts Worldwide.

Minnici, A., & Hill, D. D. (May 9, 2007). *Educational architects: Do state education agencies have the tools necessary to implement NCLB?* Washington, D.C.: Center on Education Policy.

Mitchell, A. (2010, Sep 15). Behavioural [sic] economics has yet to deliver on its promise. *Marketing*, 28-29. Retrieved from http://search.proquest.com/docview/757309665?accountid=28180

National Council on Measurement in Education (1995). *Code of professional responsibilities in educational measurement*. Washington, DC: Author.

National Forum on Education Statistics. (2006). *Forum guide to decision support systems: A resource for educators* (NFES 2006–807). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

National Forum on Education Statistics. (2010). *Traveling through time: The forum guide to longitudinal data systems. Book Two of Four: Planning and Developing an LDS* (NFES 2011–804). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

National Forum on Education Statistics. (2011). *Traveling through time: The forum guide to longitudinal data systems. Book Four of Four: Advanced LDS Usage* (NFES 2011–802). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

National Research Council (2001). *NAEP reporting practices: Investigating district-level and market-basket reporting*. Committee on NAEP Reporting Practices, Board on Testing and Assessment, Center for Education. Washington, DC: National Academy Press.

Northcentral University (2011). Informed consent checklist. *Dissertation Center*. Retrieved from http://learners.ncu.edu/ncu_diss/default.aspx

Odendahl, N. V. (2011). *Testwise: Understanding educational assessment, Volume 1*. Lanham, MD: Rowman & Littlefield Education.

Open Congress (2011). *H.R.3553 - Genetically Engineered Food Right to Know Act*. Retrieved from http://www.opencongress.org/bill/112-h3553/show

Outing, S., & Ruel, L. (2006, January 30). The Best of Eyetrack III: What we saw when we looked through their eyes. *Eyetrack III*, 1-9. The Poynter Institute. Retrieved from http://www.academia.edu/546755/The_best_of_eyetrack_III_What_we_saw_when_we_looked_through_their_eyes

Papay, J., Harvard Graduate School of Education (2007). *Aspen Institute datasheet: The teaching workforce*. Washington, DC: The Aspen Institute.

Park, V. (2008). *Beyond the numbers chase: How urban high school teachers make sense of data use*. University of Southern California. doi: AAT 3331180

Perie, M., Park, J., & Klau, K. (2007). *Key elements for educational accountability models*. Washington, DC: Council of Chief State School Officers.

Poger, S. & Bailie, F. (2006). A web-based system for course and program assessment: Prototype to product. In C. Crawford et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2006*, 120-125. Chesapeake, VA: AACE.

Prescrire. (2011). 2010 Drug packaging review: identifying problems to prevent errors. *Prescrire International, 20(117)*, 162-165. Retrieved from http://english.prescrire.org/en/115/446/48385/2408/2401/SubReportDetails.aspx

Qin, Y., Wu, M., Pan, X., Xiang, Q., Huang, J., Gu, Z., & ... Zhou, M. (2011, February 25). Reactions of Chinese adults to warning labels on cigarette packages: a survey in Jiangsu Province. *BMC Public Health, 11*(133). doi: 10.1186/1471-2458-11-133

Rankin, J. (2011, December 1). Data analysis doesn’t have to be so hard. *Ninetieth Annual California Educational Research Association (CERA) Conference*. Presentation conducted from Disney Convention Center, Anaheim, CA.

Rennie Center for Education Research and Policy. (February, 2006). *Data-driven teaching: Tools and trends*. Cambridge, MA: Rennie Center for Education Research and Policy.

Rodriguez, M. (2008, May 01). Learning how to incorporate technology in classrooms. *Inland Valley Daily Bulletin*, pp. n/a.

Sabbah, F. M. (2011). Designing more effective accountability report cards. *ProQuest Dissertations and Theses, AAT 3469488*, Retrieved from http://search.proquest.com/docview/893068662?accountid=28180

Sanchez, E.; Kline, D.; & Laird, E. (April, 2009). Data-driven districts: Building the culture and capacity to improve student achievement. *Data Quality Campaign: Using data to improve student achievement*. Retrieved from http://www.dataqualitycampaign.org/files/DQCbrief_FINAL-lowres_2_.pdf

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (1986) *Understanding regression analysis: An introductory guide*. Iowa City, IA: Sage Publications.

Schwabish, J., & Schultz, B. (2013, May 29). The Dos & Don'ts of Data Visualization. *Facebook Video Chat Interview*. Presentation conducted by All Analytics. Attended at https://www.facebook.com/AllAnalytics/app_196506863720166

Shapiro, D. L., & Smith, S. R. (2011). *Malpractice in psychology: A practical resource for clinicians*. Washington, DC: American Psychological Association

Simon, H. A. (1979). Information processing models of cognition. *Annual Review of Psychology, 30*, 363-396.

Southern Regional Education Board. (2009, March). *Guidelines for professional development of online teachers*. Atlanta, GA: Southern Regional Education Board Educational Technology Cooperative.

Stiggins, R. (2002). Assessment for learning. *Education Week, 21*(26), 30, 32-33.

Strizek, G.A., Pittsonberger, J.L., Riordan, K.E., Lyter, D.M., & Orlofsky, G.F. (2006). Characteristics of Schools, Districts, Teachers, Principals, and School Libraries in the United States: 2003-04 Schools and Staffing Survey (NCES 2006-313 Revised). *U.S. Department of Education, National Center for Education Statistics*. Washington, DC: U.S. Government Printing Office.

Supovitz, J. A., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement*. Philadelphia, PA: Consortium for Policy Research in Education.

Thaler, R. (1980). Toward a theory of consumer choice. *Journal of Economic Behavior and Organizations, 1*, 39-60.

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Tufféry, S. (2011). *Data mining and statistics for decision making* (1st ed.). West Sussex, UK: John Wiley & Sons, Ltd.

Tufte, E. R. (1997). *Visual explanations: Images and quantities, evidence and narrative*. Cheshire, CT: Graphics Press LLC.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press LLC.

Tufte, E. R. (2006). *Beautiful evidence*. Cheshire, CT: Graphics Press LLC.

Tufte, E. (2011, December 8) *Presenting Data and Information*. Presentation conducted from the Westin San Francisco Market Street, San Francisco, CA.

Underwood, J. S., Zapata-Rivera, D., & VanWinkle, W. (2008) Growing Pains: Teachers Using and Learning to Use IDMS®. *ETS Research Memorandum. RM-08-07*. Princeton, NJ: ETS.

Underwood, J. S., Zapata-Rivera, D., & VanWinkle, W. (2010). *An evidence-centered approach to using assessment data for policymakers* (ETS Research Rep. No. RR-10-03). Princeton, NJ: ETS.

U.S. Department of Education Institute of Education Sciences (IES) National Center for Education Evaluation (NCEE) and Regional Assistance (2009). *Using student achievement data to support instructional decision making*. Washington, DC: Author. (ERIC Document Reproduction Service No. ED506645)

U.S. Department of Education Institute of Education Sciences (IES) National Center for Education Evaluation (NCEE) and Regional Assistance (2010). *Error rates in measuring teacher and school performance based on student test score gains*. Washington, DC: Author. (ERIC Document Reproduction Service No. ED 511026)

U.S. Department of Education Institute of Education Sciences (IES) National Center for Education Statistics (NCES) (2012). *The condition of education 2012*. Retrieved from http://nces.ed.gov/pubs2012/2012045.pdf.

U.S. Department of Education Office of Educational Technology (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Washington, D.C.: Author.

U.S. Department of Education Office of Elementary and Secondary Education (2006). *Improving data quality for Title I standards, assessments, and accountability*

reporting: Guidelines for states, LEAs, and schools [non-regulatory guidance]. DTI Associates, A Haverstick Company, under U.S. Department of Education Contract No. ED-01-CO-0066/0009. Retrieved from http://www.google.com/url?sa=t&rct=j&q=%22guidelines%20for%20states%2C%20leas%2C%20and%20schools%22&source=web&cd=1&sqi=2&ved=0CCEQFjAA&url=http%3A%2F%2Fwww2.ed.gov%2Fpolicy%2Felsec%2Fguid%2Fstandardsassessment%2Fnclbdataguidance.doc&ei=NHKgTtKaPI-asgLYl9C_BQ&usg=AFQjCNG1gy2oDF5QzmGMtT7LwOJu_WWHXQ.

U.S. Department of Education Office of Planning, Evaluation and Policy Development (2008). *Teachers' use of student data systems to improve instruction: 2005 to 2007*. Washington, D.C.: Author.

U.S. Department of Education Office of Planning, Evaluation and Policy Development (2009). *Implementing data-informed decision making in schools: Teacher access, supports and use*. United States Department of Education (ERIC Document Reproduction Service No. ED504191)

U.S. Department of Education Office of Planning, Evaluation and Policy Development (2010). *Use of education data at the local level from accountability to instructional improvement*. United States Department of Education (ERIC Document Reproduction Service No. ED511656)

U.S. Department of Education Office of Planning, Evaluation and Policy Development (2011). *Teachers' ability to use data to inform instruction: Challenges and supports*. United States Department of Education (ERIC Document Reproduction Service No. ED516494)

U.S. Department of Labor, U.S. Bureau of Labor Statistics (2011, July). *Highlights of women's earnings in 2010: Report 1031*. Retrieved from http://www.bls.gov/cps/cpswom2010.pdf

U.S. Government Accountability Office. (2009, July). *Assessing the reliability of computer-processed data, external version I, report #GAO-09-680G*. Washington, DC: Author.

Van der Meij, H. (2008). Designing for user cognition and affect in a manual. Should there be special support for the latter? *Learning & Instruction, 18(1)*, 18-29.

VanWinkle, W., Vezzu, M., & Zapata-Rivera, D. (2011). *Question-based reports for policymakers* (ETS Research Memorandum No. RM-11-16). Princeton, NJ: ETS.

Vogt, W. P. (2006). *Quantitative research methods for professionals in education and other fields*. Boston, MA: Pearson Education.

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher, 21*(1), 14-23.

Wainer, H. (1997). *Visual Revelations: Graphical tales of fate and deception from Napoleon Bonaparte to Ross Perot.* New York, NY: Copernicus.

Watanabe, R., Gilbreath, K., & Sakamoto, C. (1994). The ability of the geriatric population to read labels on over-the-counter medication containers. *Journal of the American Optometric Association, 65(1)*, 32-37.

Wayman, J. C. (2005). Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed At Risk*, 10, no. 3: 295–308.

Wayman, J. C. (2007). Student data systems for school improvement: The state of the field. In *TCEA Educational Technology Research Symposium: Vol. 1*, 156-162. Lancaster, PA: ProActive Publications.

Wayman, J. C., & Cho, V. (2009). Preparing educators to effectively use student data systems. In T. J. Kowalski & T. J. Lasley II (Eds.), *Handbook of data-based decision making in education*, 89-104. New York, NY: Routledge.

Wayman, J. C., Cho, V., & Shaw, S. M. (2009, December). *First-year results from an efficacy study of the Acuity data system.* Paper presented at the Twenty-fourth Annual Texas Assessment Conference, Austin, TX.

Wayman, J. C., Snodgrass Rangel, V. W., Jimerson, J. B., & Cho, V. (2010). *Improving data use in NISD: Becoming a data-informed district*. Austin, TX: The University of Texas at Austin.

Wayman, J. C., & Stringfield, S. (2006). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education, 112(4)*, 549–571.

Williams, B. S. (2010). Heuristics and biases in military decision making. *Military Review, 90*(5), 40-52. Retrieved from http://search.proquest.com/docview/757053979?accountid=28180

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement, 19*(3), 239–259.

Wood, L. (2012, Sep 10). Research and markets: Patient engagement in the patient-centered medical home: A continuum approach. *M2 Presswire*, pp. n/a. Retrieved from http://search.proquest.com/docview/1038582946?accountid=28180

Zapata-Rivera, D., & VanWinkle, W. (2010). A research-based approach to designing and evaluating score reports for teachers (*ETS Research Memorandum No. RM-10-01*). Princeton, NJ: ETS.

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York, NY: Plenum Press.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (1999). *At or above proficient: The reporting of NAEP results in the Internet age*. *Center for Educational Assessment Report No. 710*. Amherst, MA: University of Massachusetts, Center for Educational Assessment.

Zwick, R., Sklar, J., Wakefield, G., Hamilton, C., Norman, A., & Folsom, D. (2008). Instructional tools in educational measurement and statistics (ITEMS) for school personnel: Evaluation of three web-based training modules. *Educational Measurement: Issues and Practice, 27*, 14–27.

Appendices

Appendix A: Standards and Codes

From 90th Annual California Educational Research Association (CERA) Conference Presentation (Rankin, 2011, pp. 39-44)

National Council on Measurement in Education:
Code of Professional Responsibilities in Educational Measurement
Responsibilities of Those Who Interpret, Use, and Communicate Assessment Results
(National Council on Measurement in Education, 1995)

Standard 6.2	Provide to those who receive assessment results information about the assessment, its purposes, its limitations, and its uses necessary for the proper interpretation of the results.
Standard 6.3	Provide to those who receive score reports an understandable written description of all reported scores, including proper interpretations and likely misinterpretations.
Standard 6.4	Communicate to appropriate audiences the results of the assessment in an understand able and timely manner, including proper interpretations and likely misinterpretations.
Standard 6.5	Evaluate and communicate the adequacy and appropriateness of any norms or standards used in the interpretation of assessment results.
Standard 6.8	Avoid making, and actively discourage others from making, inaccurate reports, unsubstantiated claims, inappropriate interpretations, or otherwise false and misleading statements about assessment results.

American Educational Research Association (AERA)
Standards for Educational & Psychological Testing
(AERA, American Psychological Association, & National Council on Measurement in Education, 1999)

Standard 5.10	"When test score information is released to students, parents, legal representatives, teachers, clients, or the media, those responsible for testing programs should provide appropriate interpretations. The interpretations should describe in simple language what the test covers, what scores mean, the precision of the scores, common misinterpretations of test scores, and how scores will be used (p. 65)."
Standard 13.1	"When educational testing programs are mandated by school, district, state, or other authorities, the ways in which tests results are intended to be used should be clearly described. It is the responsibility of those who mandate the use of the tests to monitor their impact and to identify and minimize potential negative consequences. Consequences resulting from the uses of the test, both intended and unintended, should also be examined by the test user (p. 145)."
Standard 13.9	"When tests scores are intended to be used as part of the process for making decisions for educational placement, promotion, or implementation of prescribed educational plans, empirical evidence documenting the relationship among particular test scores, the instructional programs, and desired student outcomes should be provided. When adequate empirical evidence is not available, users should be cautioned to weigh the test results accordingly in light of other relevant information about the student (p. 147)."
Standard 13.14	"In educational settings, score reports should be accompanied by a clear statement of the degree of measurement error associated with each score or classification level and information on how to interpret the scores (p. 148)."

Code of Fair Testing Practices in Education Reporting & Interpreting Test Results (Joint Committee on Testing Practices, 2004)	
Test Developers:	**Test Users:**
Test developers should report test results accurately and provide information to help test users interpret test results correctly.	Test users should report and interpret test results accurately and clearly.
C-1. Provide information to support recommended interpretations of the results, including the nature of the content, norms or comparison groups, and other technical evidence. Advise test users of the benefits and limitations of test results and their interpretation. Warn against assigning greater precision than is warranted.	C-1. Interpret the meaning of the test results, taking into account the nature of the content, norms or comparison groups, other technical evidence, and benefits and limitations of test results.
C-2. Provide guidance regarding the interpretations of results for tests administered with modifications. Inform test users of potential problems in interpreting test results when tests or test administration procedures are modified.	C-2. Interpret test results from modified test or test administration procedures in view of the impact those modifications may have had on test results.
C-3. Specify appropriate uses of test results and warn test users of potential misuses.	C-3. Avoid using tests for purposes other than those recommended by the test developer unless there is evidence to support the intended use or interpretation.
C-4. When test developers set standards, provide the rationale, procedures, and evidence for setting performance standards or passing scores. Avoid using stigmatizing labels.	C-4. Review the procedures for setting performance standards or passing scores. Avoid using stigmatizing labels.
C-5. Encourage test users to base decisions about test takers on multiple sources of appropriate information, not on a single test score.	C-5. Avoid using a single test score as the sole determinant of decisions about test takers. Interpret test scores in conjunction with other information about individuals.
C-6. Provide information to enable test users to accurately interpret and report test results for groups of test takers, including information about who were and who were not included in the different groups being compared, and information about factors that might influence the interpretation of results.	C-6. State the intended interpretation and use of test results for groups of test takers. Avoid grouping test results for purposes not specifically recommended by the test developer unless evidence is obtained to support the intended use. Report procedures that were followed…
C-7. Provide test results in a timely fashion and in a manner that is understood by the test taker.	C-7. Communicate test results in a timely fashion and in a manner that is understood by the test taker.
C-8. Provide guidance to test users about how to monitor the extent to which the test is fulfilling its intended purposes.	C-8. Develop and implement procedures for monitoring test use, including consistency with the intended purposes of the test.

Appendix B: Study Survey Pages

Note respondents only received one version of the page featuring Question 8.

10-Question Survey

Thank you so much for your time, professionalism, and feedback.

* Required

1. How long have you worked as an educator (e.g., teacher or administrator) for students under 19 years of age? *

Select the highest option applicable.

- less than 1 year
- 5 years
- 10 years
- 15 years
- 20 or more years

2. Which of the following roles best describes your current position? *

If your role is mixed, select the role requiring most of your time.

- Teacher
- Colleague Coach (e.g., Teacher on Special Assignment)
- Site/School Administrator
- District Administrator

3. How proficient are you at analyzing student performance data? *

In your opinion:

- Very proficient
- Somewhat proficient
- Not proficient
- Far from proficient

Next Steps

Please click "continue."

Continue »

Report 1

The LEFT side of your folder is labeled Report 1. Use this side's contents to answer the 2 questions below.

4. Which content cluster is most likely the School's strength? *

Base your answer on the folder's Report 1.

- Word Analysis and Vocabulary Development
- Reading Comprehension
- Literary Response and Analysis
- Written Conventions
- Writing Strategies
- Writing Applications

5. Which content cluster is most likely the School's weakness? *

Base your answer on the folder's Report 1.

- Word Analysis and Vocabulary Development
- Reading Comprehension
- Literary Response and Analysis
- Written Conventions
- Writing Strategies
- Writing Applications

Next Steps

After you are finished with both questions above, please return your report materials to the LEFT side of your folder.

After that, click "continue."

« Back   Continue »

Report 2

The RIGHT side of your folder is labeled Report 2. Use this side's contents to answer the 2 questions below.

6. Which student(s) did NOT score Proficient on the CELDT? *

Base your answer on the folder's Report 2. CHECK ALL THAT APPLY.

- [ ] Student A
- [ ] Student B
- [ ] Student C
- [ ] Student D

7. In which area(s) did at least 1 student earn a score that PREVENTED him/her from scoring Proficient on the CELDT? *

Base your answer on the folder's Report 2. CHECK ALL THAT APPLY.

- [ ] Listening
- [ ] Speaking
- [ ] Reading
- [ ] Writing
- [ ] Overall

Next Steps

After you are finished with both questions above, please return your report materials to the RIGHT side of your folder.

After that, please answer the question below.

What color is your folder? *

The cover of your report materials folder features the name of its color.

- ○ White
- ○ Yellow
- ○ Green
- ○ Blue
- ○ Purple
- ○ Red
- ○ Black

Next Steps

Please click "continue."

[« Back] [Continue »]

8. The 2 reports you just used did not offer any special assistance in analyzing the data. If they had been accompanied by text (e.g., a footer, guide, or abstract) designed to help you interpret the data, would you likely have used the added support?

- Yes – I probably would use the support.
- No – I probably would not use the support.

Next Steps

You're almost done. Please click "continue."

« Back   Continue »

8. The 2 reports you just used contained footers with analysis guidelines designed to help you. Did you read these footers before answering questions related to the reports? *

- Yes – I referred to both reports' footers.
- I referred to Report 1's footer but not Report 2's footer.
- I referred to Report 2's footer but not Report 1's footer.
- No – I did not refer to either footer.

Next Steps

You're almost done. Please click "continue."

« Back | Continue »

8. The 2 reports you just used were each accompanied by a 1-page abstract (like a reference sheet) with analysis guidelines designed to help you. Did you read these abstracts/sheets before answering questions related to the reports? *

- Yes – I referred to both reports' abstracts/sheets.
- I referred to Report 1's abstract/sheet but not Report 2's abstract/sheet.
- I referred to Report 2's abstract/sheet but not Report 1's abstract/sheet.
- No – I did not refer to either abstract/sheet.

Next Steps

You're almost done. Please click "continue."

« Back | Continue »

8. The 2 reports you just used were each accompanied by an interpretation guide (a packet) with analysis guidelines designed to help you. Did you read these guides before answering questions related to the reports? *

- ◯ Yes – I referred to both reports' guides.
- ◯ I referred to Report 1's guide but not Report 2's guide.
- ◯ I referred to Report 2's guide but not Report 1's guide.
- ◯ No – I did not refer to either guide.

Next Steps

You're almost done. Please click "continue."

« Back   Continue »

Only 2 Questions Left

9. Lots of professional development happens at school sites: for example, demonstrations to accompany textbook adoptions, meetings with colleagues to share differentiation strategies, training on how to use new software, etc. Only some professional development specifically focuses on how to analyze student data. Within the last 12 months, how many hours of professional development have you had that specifically focused on teaching you how to correctly interpret student data? *

Select the highest option applicable. Time spent analyzing student data without guidance should not be counted, nor should time spent learning technology to generate student data.

- 0 hours
- 1 hour
- 2 hours
- 5 hours
- 8 or more

10. Educational Measurement refers to the analysis of student assessment data to draw conclusions about abilities. How many graduate-level courses have you taken that were specifically dedicated to educational measurement (e.g., student performance data analysis, measurement theory, or psychometrics)? *

Select the highest option applicable.

- 0 courses
- 1 course
- 2 courses
- 3 courses
- 4 or more

Next Steps

After you are finished with both questions above, please click "submit" and then raise your folder in the air for Jenny to collect. Please keep your computer on after clicking "submit."

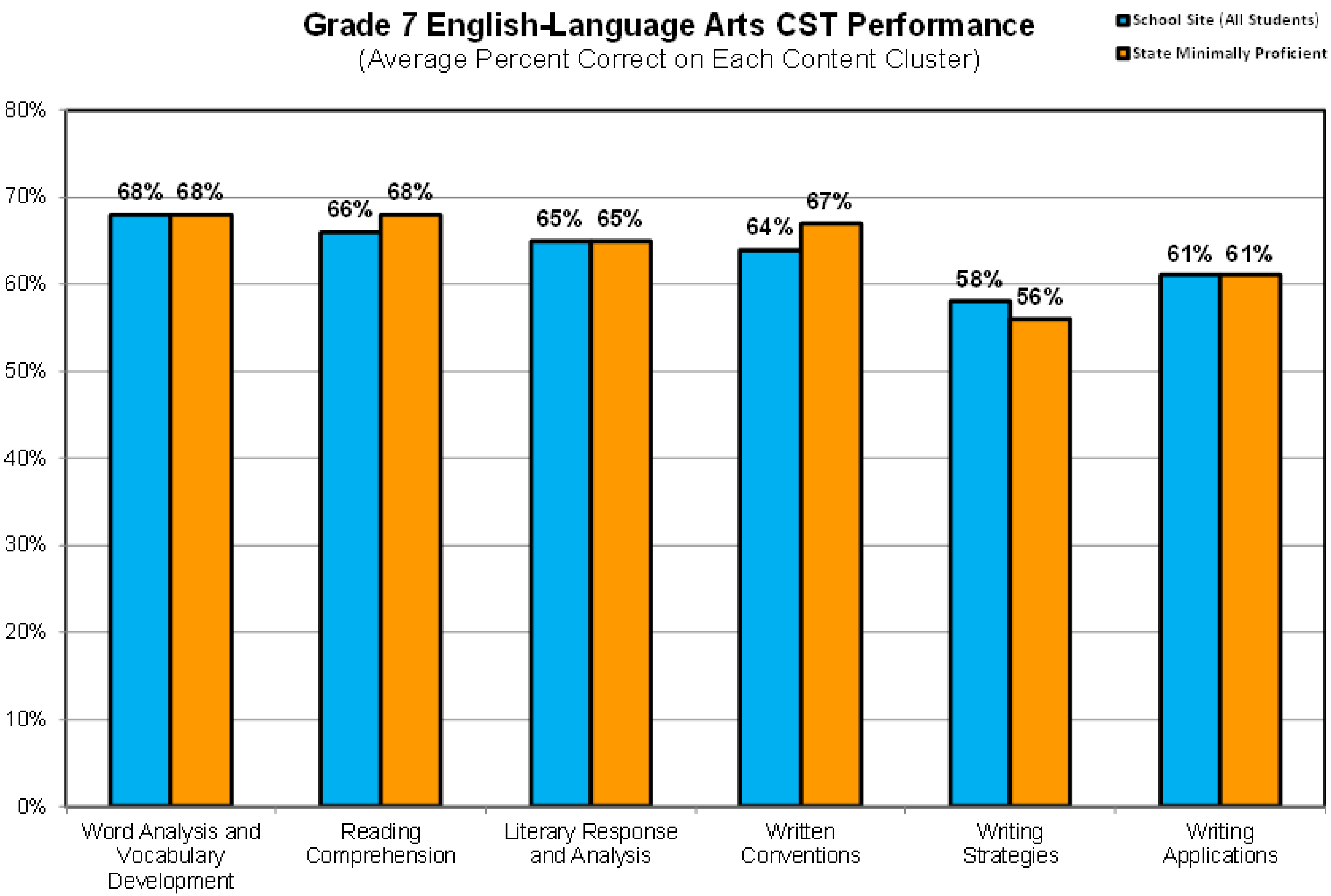
« Back   Submit

Appendix C: Handouts Used in Study (Color Format Is Pertinent to Study)

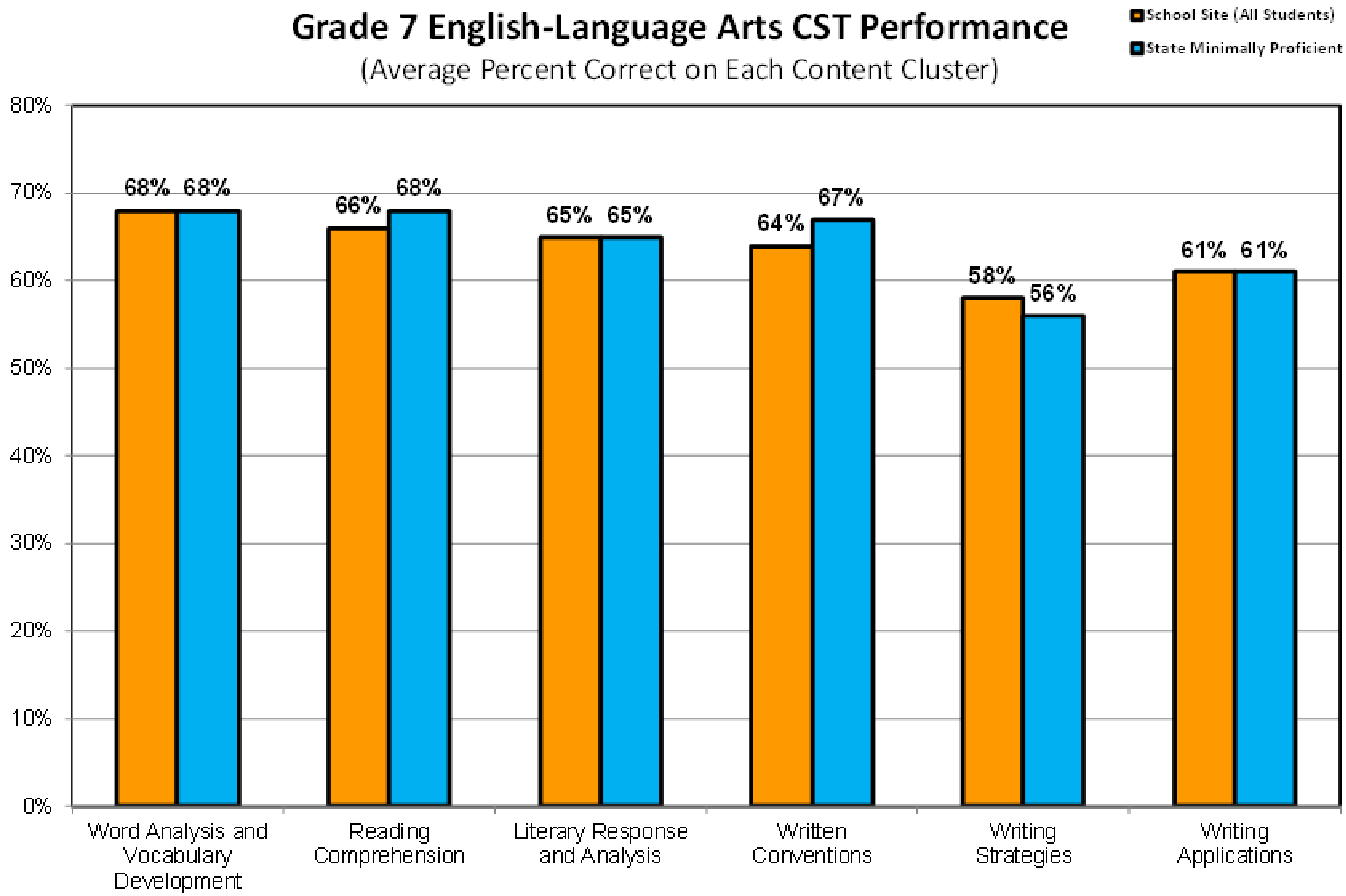
The succeeding 20 pages contain the following handouts, which are followed by depictions of how they were distributed to participants:

Item	Included in Folder (Scenario #)	
Report 1 with No Footer (Plain Report)	White (1) Purple (4) Blue (5)	Black (6) Red (7)
Report 1 with Footer A (Shorter)	Green (2)	
Report 1 with Footer B (Longer)	Yellow (3)	
Abstract 1A (Less Dense)	Purple (5)	
Abstract 1B (Denser)	Blue (5)	
Interpretation Guide 1A (2 Pages)	Black (6)	
Interpretation Guide 1B (3 Pages)	Red (7)	
Report 2 with No Footer (Plain Report)	White (1) Purple (4) Blue (5)	Black (6) Red (7)
Report 2 with Footer A (Shorter)	Green (2)	
Report 2 with Footer B (Longer)	Yellow (3)	
Abstract 2A (Less Dense)	Purple (5)	
Abstract 2B (Denser)	Blue (5)	
Interpretation Guide 2A (2 Pages)	Black (6)	
Interpretation Guide 2B (3 Pages)	Red (7)	

REPORT 1


Grade 7 English-Language Arts CST Performance
(Average Percent Correct on Each Content Cluster)
School Site (All Students)
State Minimally Proficient
80%
70%
60%
50%
40%
30%
20%
10%
0%
68% 68%
66% 68%
65% 65%
64% 67%
58% 56%
61% 61%
Word Analysis and Vocabulary Development
Reading Comprehension
Literary Response and Analysis
Written Conventions
Writing Strategies
Writing Applications

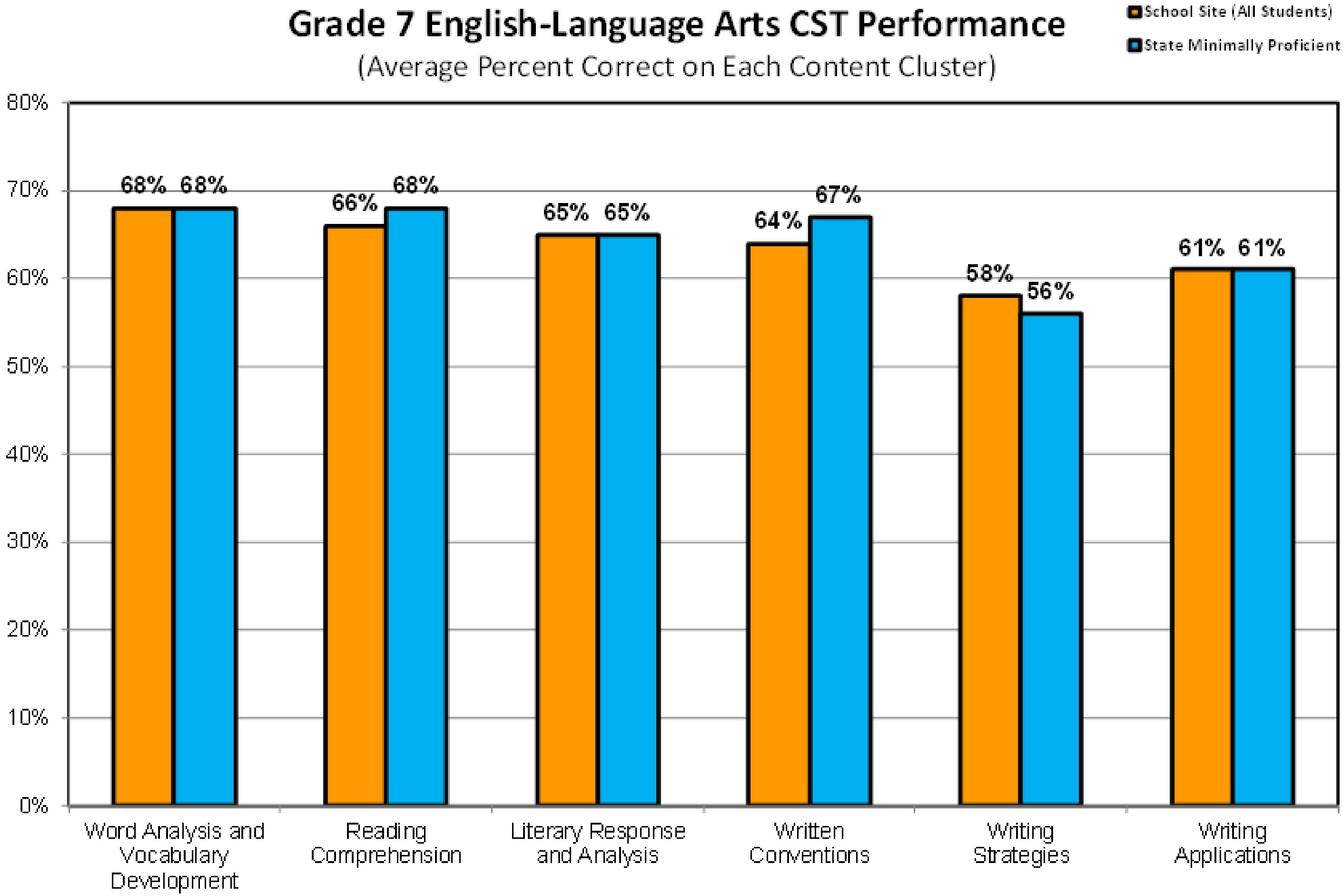
REPORT 1



Warning: Clusters vary in difficulty, so the Site's highest % correct is <u>not</u> necessarily a strength.

What to Do: <u>Site % – State Minimally Proficient % = #</u> (highest # could be Site strength, lowest # could be Site weaknesses).

REPORT 1



Clusters vary in difficulty, so the Site's highest % correct is not necessarily a strength.
Compare the Site % to the State Minimally Proficient % (i.e., look at the degree to which the Site *beat* the SMP).
Site % - SMP % = # (cluster with highest difference could be Site strength, lowest difference could be Site weaknesses).

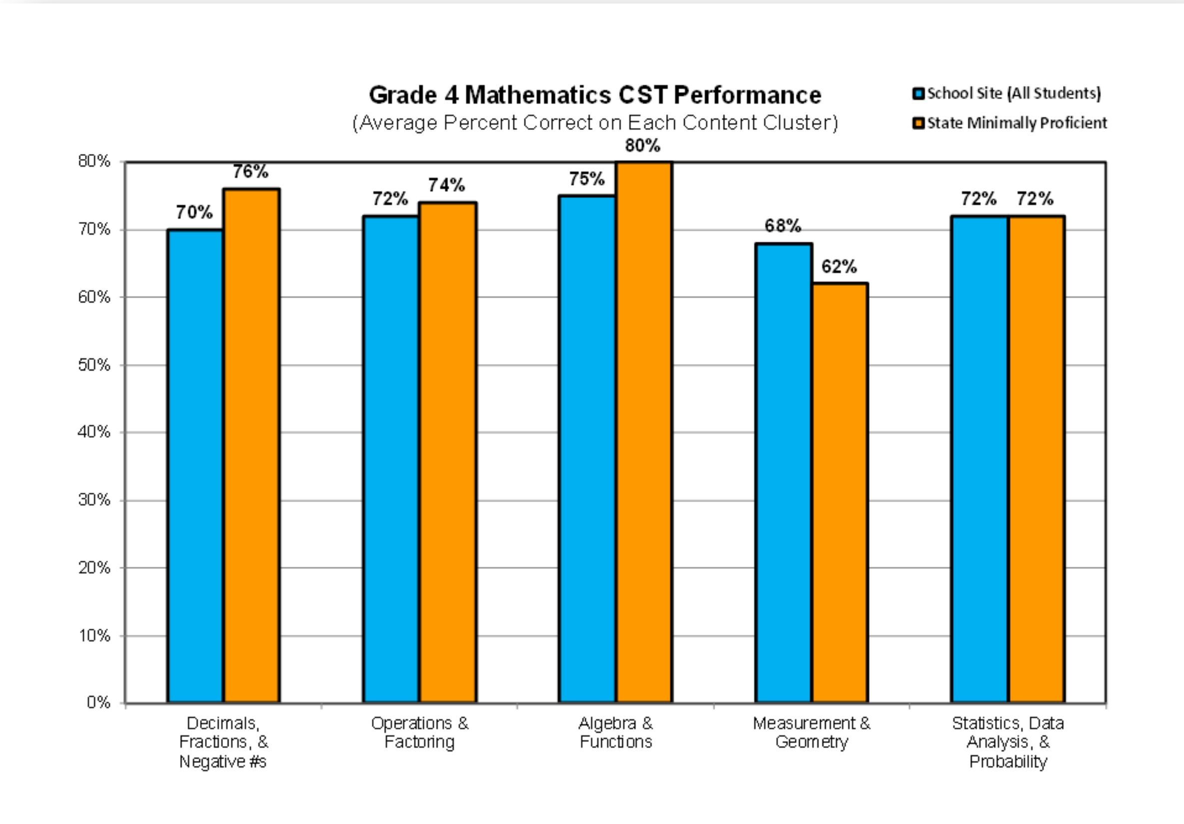
CST Performance Report
Abstract

This page provides an abstract for the *CST Performance* report, which shows a school site's performance on California Standards Test (CST) content clusters in relation to the state's performance (scores of students statewide who scored *Proficient* on the CST).

Focus **What data is reported?**

Students' average % correct when answering questions aligned to each CST content cluster is displayed for:

- a school site
- the State Minimally Proficient (meaning all students in California who scored the minimum scale score needed – 350 – to be considered *Proficient* on this CST)



Warning **What do many educators misunderstand?**

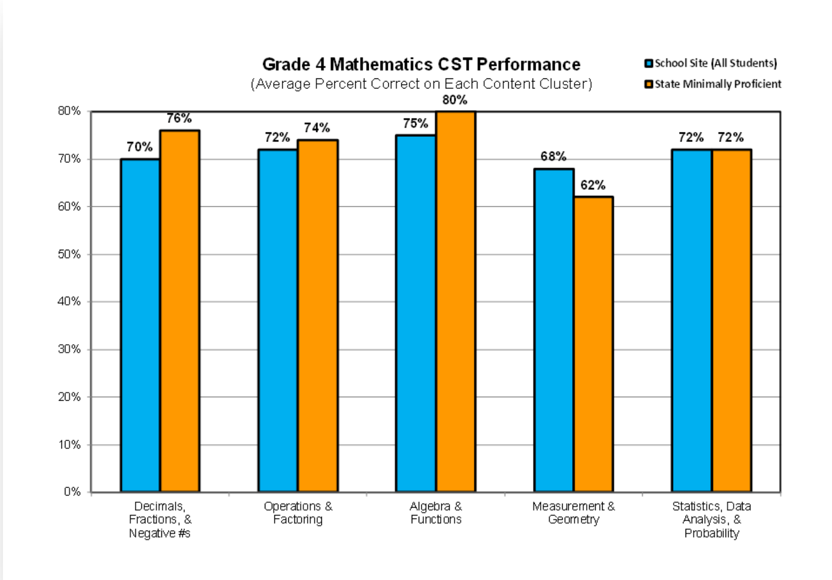
Content clusters vary in difficulty, so a site's highest % correct for a cluster does not necessarily indicate its strength, and its lowest % correct for a cluster is not necessarily its weakness. For each cluster, compare the Site % to the State Minimally Proficient % (i.e., *look at the degree to which the Site beat the State Minimally Proficient*). Use this formula:

<u>School Site %</u> – <u>State Minimally Proficient %</u> = <u>#</u>

The cluster with the highest difference (highest # from above formula) could be a Site strength, and the cluster with the lowest difference (lowest # from above formula) could be a Site weaknesses.

CST Performance Report
Abstract

This page provides an abstract for the *CST Performance* report, which shows a school site's performance on California Standards Test (CST) content clusters in relation to the state's performance (scores of students statewide who scored *Proficient* on the CST).



Purpose **What are some questions this report will help answer?**

- What are possible weaknesses for my school site (in a grade and subject area)?
- What are possible strengths for my school site (in a grade and subject area)?
- Which content clusters were assessed with the hardest questions on this CST?
- Which content clusters were assessed with the easiest questions on this CST?

Focus **Who is the intended audience?**

Teachers and administrators

What data is reported?

Students' average % correct when answering questions aligned to each CST content cluster is displayed for:

- a school site
- the State Minimally Proficient (meaning all students in California who scored the minimum scale score needed – 350 – to be considered *Proficient* on this CST)

How is the data reported?

The school site is graphed in blue, and the State Minimally Proficient is graphed in orange.

Warning **What do many educators misunderstand?**

Content clusters vary in difficulty, so a site's highest % correct for a cluster does not necessarily indicate its strength, and its lowest % correct for a cluster is not necessarily its weakness. For each cluster, compare the Site % to the State Minimally Proficient % (i.e., *look at the degree to which the Site beat the State Minimally Proficient*). Use this formula:

School Site % – **State Minimally Proficient %** = **#**

The cluster with the highest difference (highest # from above formula) could be a Site strength, and the cluster with the lowest difference (lowest # from above formula) could be a Site weaknesses.

CST Performance Report
Interpretation Guide

The *CST Performance* report shows a school site's performance on California Standards Test (CST) content clusters in relation to the state's performance.

Warning **What do many educators misunderstand?**

Content clusters vary in difficulty, so a site's highest % correct for a cluster does not necessarily indicate its strength, and its lowest % correct for a cluster is not necessarily its weakness. For each cluster, compare the Site % to the State Minimally Proficient % (i.e., *look at the degree to which the Site beat the State Minimally Proficient*). Use this formula:

<u>School Site %</u> – <u>State Minimally Proficient %</u> = <u>#</u>

The cluster with the highest difference (highest # from above formula) could be a Site strength, and the cluster with the lowest difference (lowest # from above formula) could be a Site weaknesses.

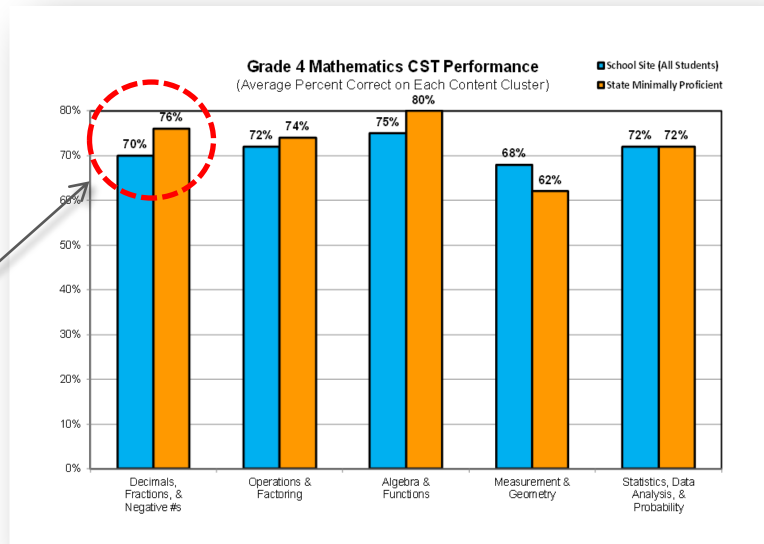
Essential Questions **What are possible weaknesses for my school site (in a grade and subject area)?**

Determine the cluster in which you most lagged behind the State Minimally Proficient's (SMP's) students (or beat them to the least degree). Since clusters vary in difficulty, SMP %s account for how easy or hard the clusters were. Use this formula:

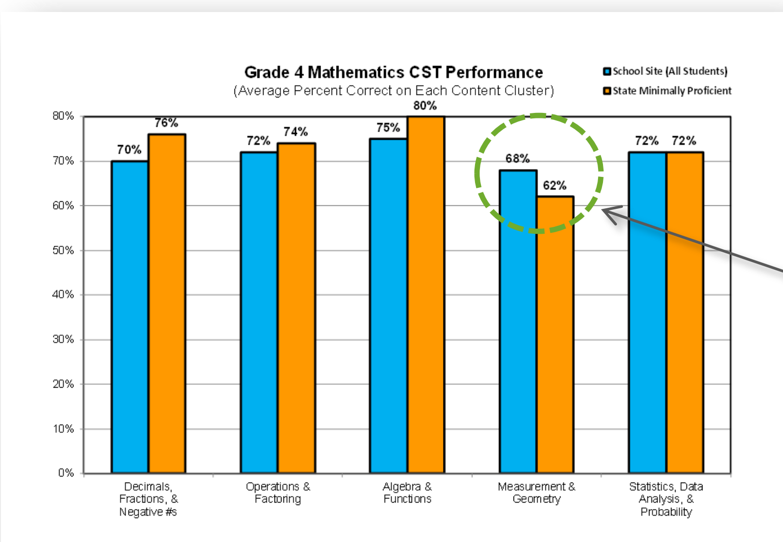
<u>School %</u> – <u>SMP %</u> = <u>#</u>

Example: For the *Decimals* cluster:

<u>School 70%</u> – <u>SMP 76%</u> = <u>–6</u>



More than for any other cluster, Site did most poorly on the *Decimals* cluster (because of how Site compared to SMP). The *Decimals* cluster is most likely Site's weakness, even though the Site's 70% for *Decimals* was not its lowest %.



What are possible strengths for my school site (in a grade and subject area)?

Determine the cluster in which you beat the State Minimally Proficient's (SMP's) students to the greatest degree. Since clusters vary in difficulty, SMP %s account for how easy or hard the clusters were. Use this formula:

<u>School %</u> – <u>SMP %</u> = <u>#</u>

Example: For the *Measurement* cluster:

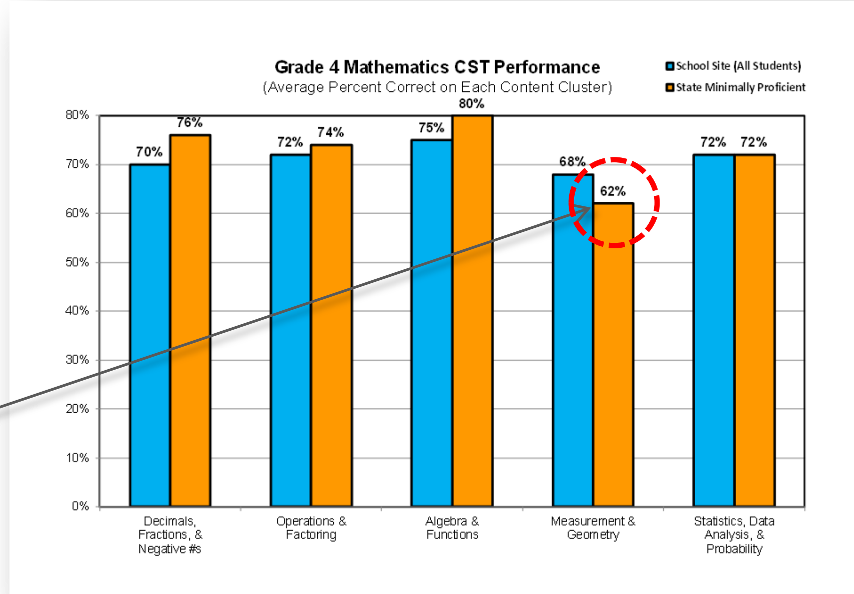
<u>School 68%</u> – <u>SMP 62%</u> = <u>+6</u>

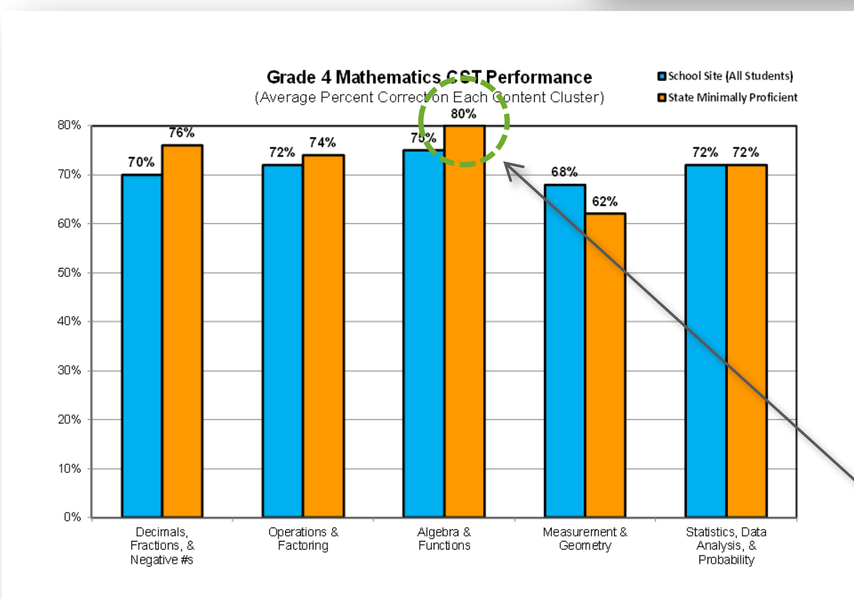
More than for any other cluster, Site performed best on the *Measurement* cluster (because of how Site compared to SMP).
The *Measurement* cluster is most likely Site's strength, even though the Site's 68% for *Measurement* was not its highest %.

Which content clusters were assessed with the hardest questions on this CST?

Find the State Minimally Proficient (SMP) lowest %. Since SMP %s are the average % of questions answered correctly by all students in California who scored the minimum scale score needed – 350 – to be considered *Proficient* on this CST, clusters they struggled with the most had the hardest questions.

Example: SMP's 62% in *Measurement* is lower than the 76%, 74%, 80%, and 72% SMP earned in the other clusters. Thus the *Measurement* cluster was likely assessed with the hardest questions.





Which content clusters were assessed with the easiest questions on this CST?

Find the State Minimally Proficient (SMP) highest %. Clusters that SMP had the easiest time with had the easiest questions.

Example: SMP's 80% in *Algebra* is higher than the 76%, 74%, 62%, and 72% SMP earned in the other clusters. Thus the *Algebra* cluster was likely assessed with the easiest questions.

More Info

Where can I find more info on the CST and its proper analysis?

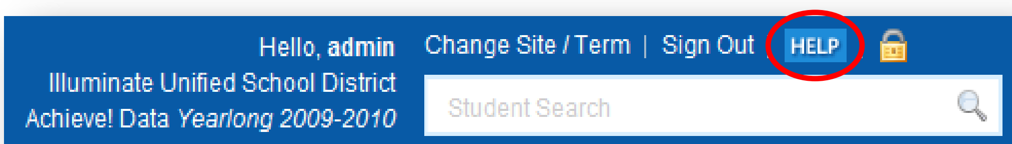
Reference Chapter 1 of the *California Standardized Testing and Reporting (STAR) Post-Test Guide* at **http://www.startest.org/archive.html**.

Where can I find more info on analyzing CST content clusters?

Visit the Help system's *Data Analysis* manual.

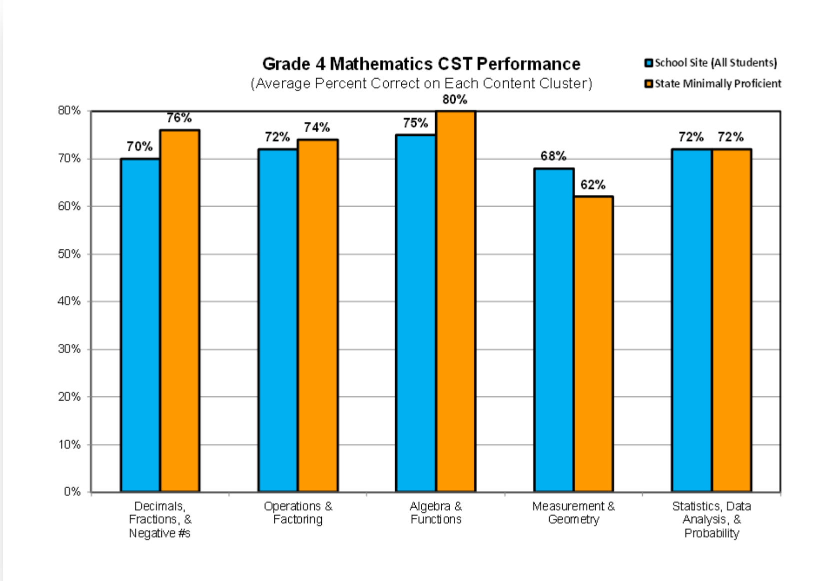
Where can I learn how to generate this report in my data system?

Visit the Help system's *Reports* manual.



CST Performance Report
Interpretation Guide

This 3-page guide explains the *CST Performance* report, which shows a school site's performance on California Standards Test (CST) content clusters in relation to the state's performance (scores of students statewide who scored *Proficient* on the CST).



Purpose **What are some questions this report will help answer?**

- What are possible weaknesses for my school site (in a grade and subject area)?
- What are possible strengths for my school site (in a grade and subject area)?
- Which content clusters were assessed with the hardest questions on this CST?
- Which content clusters were assessed with the easiest questions on this CST?

Focus **Who is the intended audience?**

Teachers and administrators

What data is reported?

Students' average % correct when answering questions aligned to each CST content cluster is displayed for:

- a school site
- the State Minimally Proficient (meaning all students in California who scored the minimum scale score needed – 350 – to be considered *Proficient* on this CST)

How is the data reported?

The school site is graphed in blue, and the State Minimally Proficient is graphed in orange.

Warning **What do many educators misunderstand?**

Content clusters vary in difficulty, so a site's highest % correct for a cluster does not necessarily indicate its strength, and its lowest % correct for a cluster is not necessarily its weakness. For each cluster, compare the Site % to the State Minimally Proficient % (i.e., *look at the degree to which the Site beat the State Minimally Proficient*). Use this formula:

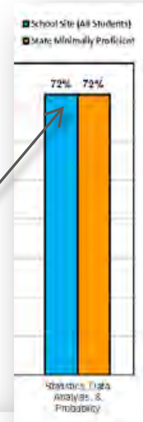
School Site % – **State Minimally Proficient %** = **#**

The cluster with the highest difference (highest # from above formula) could be a Site strength, and the cluster with the lowest difference (lowest # from above formula) could be a Site weaknesses.

Instructions

How do I read the report?

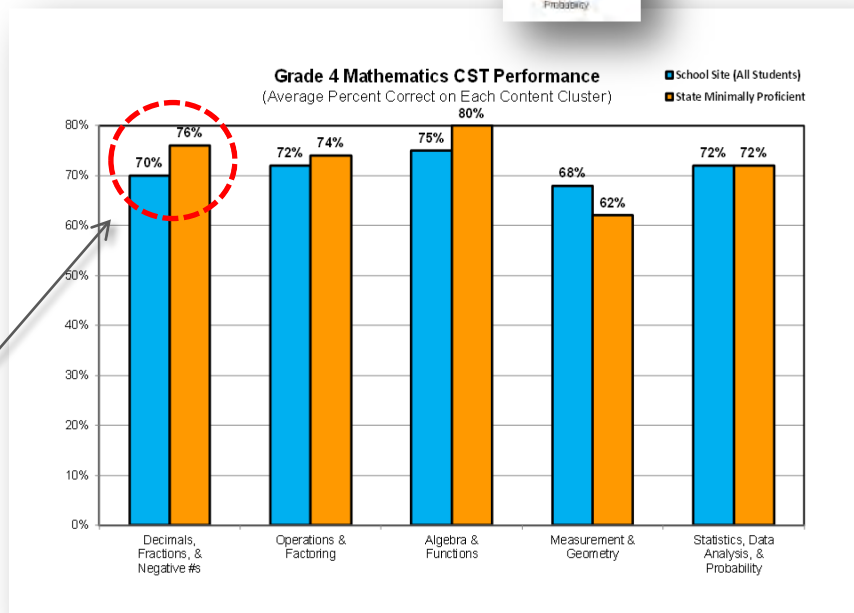
The bars show you the % of questions students answered correctly when answering questions aligned to each CST content cluster. %s above blue bars are results of students at the School Site, and %s above orange bars are results of students statewide who scored the minimum scale score needed (350) to be considered *Proficient* on this CST.



Example: The State Minimally Proficient students *and* the School Site's students both answered 72% of Qs correctly in this CST's *Statistics* cluster.

Essential Questions

What are possible weaknesses for my school site (in a grade and subject area)?



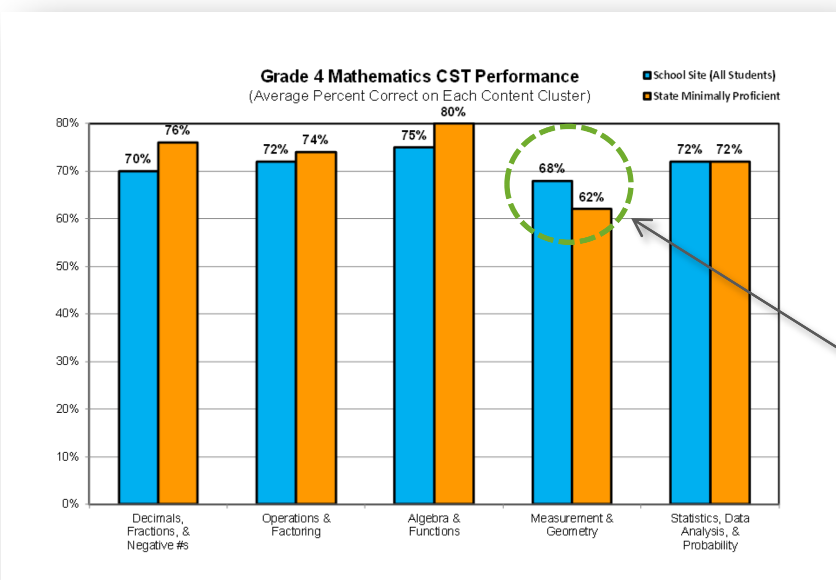
Determine the cluster in which you most lagged behind the State Minimally Proficient's (SMP's) students (or beat them to the least degree). Since clusters vary in difficulty, SMP %s account for how easy or hard the clusters were. Use this formula:

School % – **SMP %** = **#**

Example: For the *Decimals* cluster:

School 70% – SMP 76% = –6

More than for any other cluster, Site did most poorly on the *Decimals* cluster (because of how Site compared to SMP). The *Decimals* cluster is most likely Site's weakness, even though the Site's 70% for *Decimals* was not its lowest %.



What are possible strengths for my school site (in a grade and subject area)?

Determine the cluster in which you beat the State Minimally Proficient's (SMP's) students to the greatest degree. Since clusters vary in difficulty, SMP %s account for how easy or hard the clusters were. Use this formula:

School % – **SMP %** = **#**

Example: For the *Measurement* cluster:

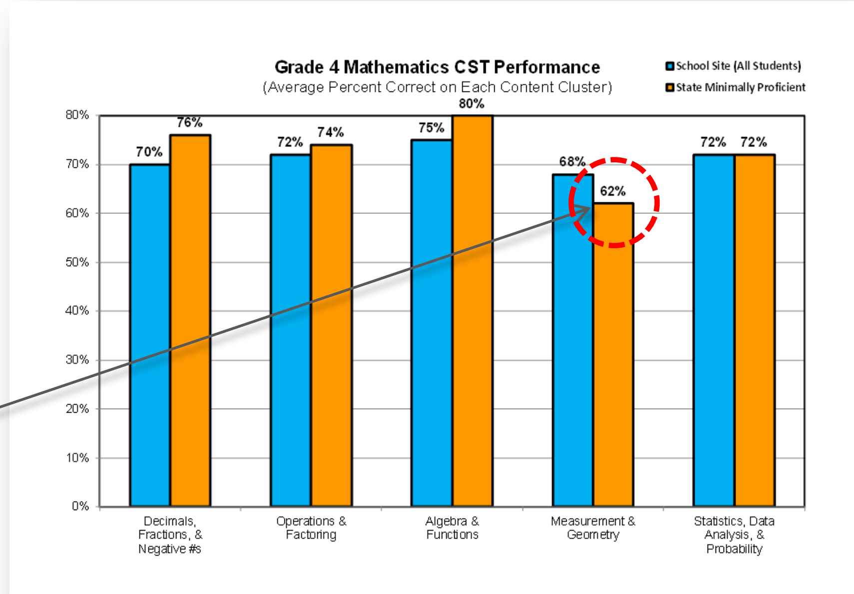
School 68% – SMP 62% = +6

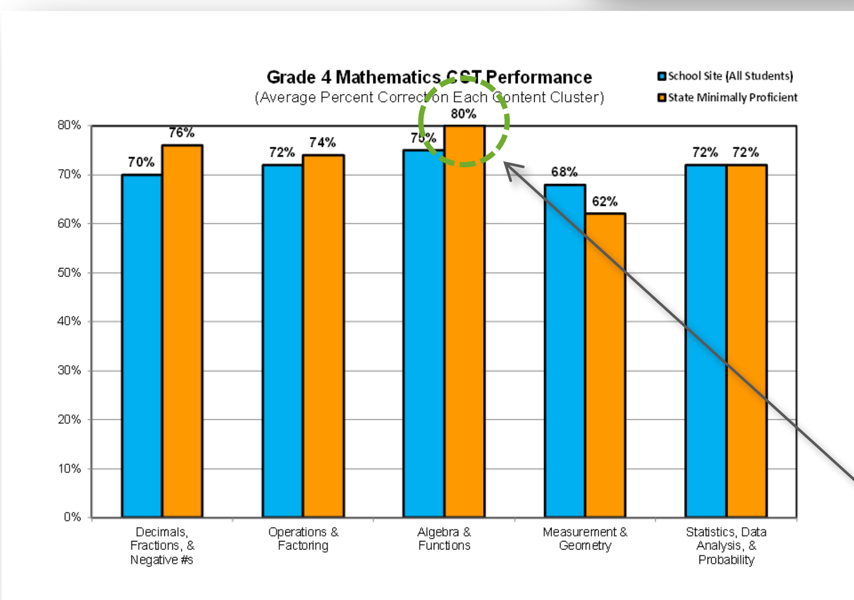
More than for any other cluster, Site performed best on the *Measurement* cluster (because of how Site compared to SMP).
The *Measurement* cluster is most likely Site's strength, even though the Site's 68% for *Measurement* was not its highest %.

Which content clusters were assessed with the hardest questions on this CST?

Find the State Minimally Proficient (SMP) lowest %. Since SMP %s are the average % of questions answered correctly by all students in California who scored the minimum scale score needed – 350 – to be considered *Proficient* on this CST, clusters they struggled with the most had the hardest questions.

Example: SMP's 62% in *Measurement* is lower than the 76%, 74%, 80%, and 72% SMP earned in the other clusters. Thus the *Measurement* cluster was likely assessed with the hardest questions.





Which content clusters were assessed with the easiest questions on this CST?

Find the State Minimally Proficient (SMP) highest %. Clusters that SMP had the easiest time with had the easiest questions.

Example: SMP's 80% in *Algebra* is higher than the 76%, 74%, 62%, and 72% SMP earned in the other clusters. Thus the *Algebra* cluster was likely assessed with the easiest questions.

More Info **Where can I find more info on the CST and its proper analysis?**

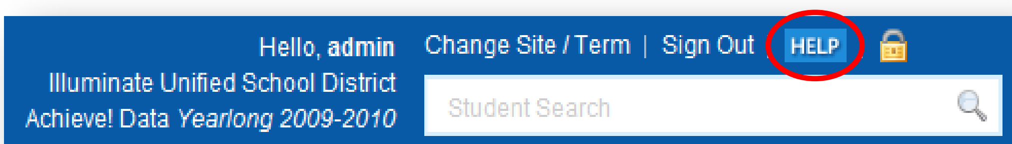
Reference Chapter 1 of the *California Standardized Testing and Reporting (STAR) Post-Test Guide* at **http://www.startest.org/archive.html**.

Where can I find more info on analyzing CST content clusters?

Visit the Help system's *Data Analysis* manual.

Where can I learn how to generate this report in my data system?

Visit the Help system's *Reports* manual.



REPORT 2

Students' CELDT Performance

(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Student A	2	3	3	4	5	4
Student B	7	3	3	4	4	3
Student C	5	4	5	4	5	4
Student D	11	4	2	5	5	5
Average		3.5	3.3	4.3	4.8	4.0

Students' CELDT Performance

(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Student A	2	3	3	4	5	4
Student B	7	3	3	4	4	3
Student C	5	4	5	4	5	4
Student D	11	4	2	5	5	5
Average		3.5	3.3	4.3	4.8	4.0

Warning: "Overall" is not the only score that determines CELDT proficiency.

What to Do: Consider a student CELDT *Proficient* only with both:
- ✓ 4 or above *Overall* , &
- ✓ 3 or above in every domain

Students' CELDT Performance

(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Student A	2	3	3	4	5	4
Student B	7	3	3	4	4	3
Student C	5	4	5	4	5	4
Student D	11	4	2	5	5	5
Average		3.5	3.3	4.3	4.8	4.0

The student's "Overall" score is not the only score that determines CELDT proficiency.
A student is *Proficient* on the CELDT only if earning both of these:
- performance level 4 or above *Overall*, &
- performance level 3 or above in every domain

Students’ CELDT Performance
Abstract

This page provides an abstract for the *Students’ CELDT Performance* report, which shows English Learners’ scores on the California English Language Development Test (CELDT), which determines which students should be considered for reclassification as Fluent English Proficient (RFEP).

Focus **What data is reported?**

Each English Learner who took the CELDT is listed with grade level, proficiency level for each domain, and *Overall* proficiency level.

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	2	2	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

Warning **What do many educators misunderstand?**

The *Overall* score does <u>not</u>, alone, determine CELDT proficiency. A Grade 2-12 student is *Proficient* on the CELDT <u>only</u> if earning <u>both</u> of these:

- performance level 4 or above *Overall*
- performance level 3 or above in every domain

Kindergarten and Grade 1 students only have to meet these criteria for Listening, Speaking, and Overall in order to score *Proficient*.

Students' CELDT Performance
Abstract

This page provides an abstract for the *Students' CELDT Performance* report, which shows English Learners' scores on the California English Language Development Test (CELDT), which determines which students should be considered for reclassification as Fluent English Proficient (RFEP).

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	2	2	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

Purpose What are some questions this report will help answer?

- Which students scored *Proficient* on the CELDT?
- Which scores prevented students from earning Proficiency?
- How did this class or program of students perform on the CELDT and in each of its domains?

Focus Who is the intended audience?

Teachers, administrators, and EL coordinators

What data is reported?

Each English Learner who took the CELDT is listed with grade level, proficiency level for each domain, and *Overall* proficiency level.

How is the data reported?

Students in a class or program are listed with their scores. A final row averages all the scores in each domain and *Overall*.

Warning What do many educators misunderstand?

The *Overall* score does not, alone, determine CELDT proficiency. A Grade 2-12 student is *Proficient* on the CELDT only if earning both of these:

- performance level 4 or above *Overall*
- performance level 3 or above in every domain

Kindergarten and Grade 1 students only have to meet these criteria for Listening, Speaking, and Overall in order to score *Proficient*.

Students' CELDT Performance
Interpretation Guide

The *Students' CELDT Performance* report shows English Learners' scores on the CELDT, a test that determines which students should be considered for reclassification.

Warning **What do many educators misunderstand?**

The *Overall* score does not, alone, determine CELDT proficiency. A Grade 2-12 student is *Proficient* on the CELDT only if earning both of these:

- performance level 4 or above *Overall*
- performance level 3 or above in every domain

Kindergarten and Grade 1 students only have to meet these criteria for Listening, Speaking, and Overall in order to score *Proficient*.

Essential Questions **Which students scored *Proficient* on the CELDT?**

To determine who scored Proficient, you must consider the *Overall* score and the domain scores.

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	2	2	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

Grades 2-12

A student is *Proficient* only if earning both of these*:

- 4 or above *Overall*
- 3 or above in every domain

Example: Ashley is not *Proficient* because of her 2 in *Reading*.

Example: Victor is not *Proficient* because of his 3 *Overall*.

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	2	2	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

Grades K-1

A K-1 student is *Proficient* only if earning both of these:

- 4 or above *Overall*
- 3 or above in *Listening*
- 3 or above in *Speaking*

Example: Cho is *Proficient* because of her 5s ("3 or above") in *Listening* and *Speaking* and her 4 ("4 or above") *Overall*. Because she is in Kindergarten her 2s aren't considered.

*K-1 Grade students are an exception to the above rules in that only their *Listening*, *Speaking*, and *Overall* scores are considered when determining Proficiency.

Which scores prevented students from earning Proficiency?

Find every 1 or 2 in the Domain area (remember to ignore K-1 students' *Reading* and *Writing* scores).

Find every 1, 2, and 3 in the *Overall* area.

Example: All but the *Speaking* domain caused students in this program to not earn Proficiency.

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	2	2	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

How did this class or program of students perform on the CELDT and in each of its domains?

Reference the bottom row to view class or program averages.

Average	3.8	4.3	2.3	3.0	3.3

Example: This program's average of 4.3 (for *Speaking*) was highest for all the domains, whereas 2.3 (for *Reading*) was its lowest. This program's *Overall* average was 3.3.

More Info **Where can I find more info on the CELDT?**

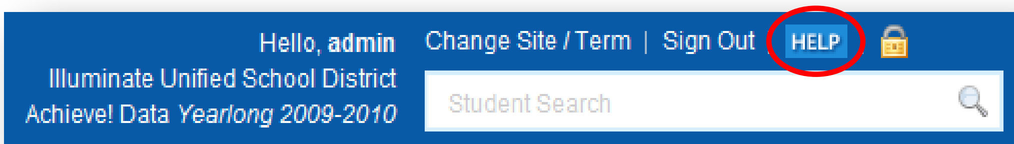
Visit http://www.cde.ca.gov/ta/tg/el/ for resources.

Where can I find more info on analyzing CELDT performance?

Visit the Help system's *Data Analysis* manual**.**

Where can I learn how to generate this report in my data system?

Visit the Help system's *Reports* manual**.**



Who takes the CELDT and when?

All students whose home language is not English must test within 30 calendar days of enrolling in a California public school to determine classification as Fluent-English Proficient (FEP) or English Learner (EL). ELs must test every year thereafter until they are Reclassified as Fluent-English Proficient (R-FEP).

What do the performance levels mean?

1 = Beginning, 2 = Early Intermediate, 3 = Intermediate, 4 = Early Advanced, 5 = Advanced

Students' CELDT Performance
Interpretation Guide

This 3-page guide explains the *Students' CELDT Performance* report, which shows English Learners' scores on the California English Language Development Test (CELDT), which determines which students should be considered for reclassification as Fluent English Proficient (RFEP).

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	2	2	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

Purpose **What are some questions this report will help answer?**

- Which students scored *Proficient* on the CELDT?
- Which scores prevented students from earning Proficiency?
- How did this class or program of students perform on the CELDT and in each of its domains?

Focus **Who is the intended audience?**

Teachers, administrators, and EL coordinators

What data is reported?

Each English Learner who took the CELDT is listed with grade level, proficiency level for each domain, and *Overall* proficiency level.

How is the data reported?

Students in a class or program are listed with their scores. A final row averages all the scores in each domain and *Overall*.

Warning **What do many educators misunderstand?**

The *Overall* score does not, alone, determine CELDT proficiency. A Grade 2-12 student is *Proficient* on the CELDT only if earning both of these:

- performance level 4 or above *Overall*
- performance level 3 or above in every domain

Kindergarten and Grade 1 students only have to meet these criteria for Listening, Speaking, and Overall in order to score *Proficient*.

Instructions

How do I read the report?

Each English Learner has his or her own row of scores. The 1st 4 of these scores are for domains, which are categories of English-Language Development (ELD) standards on which the test assesses students. The final score summarizes the student's *Overall* CELDT performance. However, this *Overall* score does not, alone, determine CELDT proficiency.

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	2	2	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

Essential Questions

Which students scored *Proficient* on the CELDT?

To determine who scored Proficient, you must consider the *Overall* score <u>and</u> the domain scores.

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	2	2	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

<u>Grades 2-12</u>

A student is *Proficient* <u>only</u> if earning <u>both</u> of these:

- 4 or above *Overall*
- 3 or above in every domain

Example: Ashley is <u>not</u> *Proficient* because of her 2 in *Reading*.

Example: Victor is <u>not</u> *Proficient* because of his 3 *Overall*.

Kindergarten and Grade 1 students are an exception to the above rules in that only their *Listening*, *Speaking*, and *Overall* scores are considered when determining Proficiency.

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	~~2~~	~~2~~	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

<u>Grades K-1</u>

A K-1 student is *Proficient* <u>only</u> if earning <u>both</u> of these:

- 4 or above *Overall*
- 3 or above in *Listening*
- 3 or above in *Speaking*

Example: Cho <u>is</u> *Proficient* because of her 5s ("3 or above") in *Listening* and *Speaking* and her 4 ("4 or above") *Overall*. Because she is in Kindergarten her 2s aren't considered.

Which scores prevented students from earning Proficiency?

Find every 1 or 2 in the Domain area (remember to ignore K-1 students' *Reading* and *Writing* scores).

Find every 1, 2, and 3 in the *Overall* area.

Example: All but the *Speaking* domain caused students in this program to not earn Proficiency.

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Ashley Garcia	4	5	5	2	4	4
Victor Jung	11	3	4	3	4	3
Cho McDonald	Kindergarten	5	5	2	2	4
Jose Patel	8	2	3	2	2	2
Average		3.8	4.3	2.3	3.0	3.3

How did this class or program of students perform on the CELDT and in each of its domains?

Reference the bottom row to view class or program averages.

Average	3.8	4.3	2.3	3.0	3.3

Example: This program's average of 4.3 (for *Speaking*) was highest for all the domains, whereas 2.3 (for *Reading*) was its lowest. This program's *Overall* average was 3.3.

More Info **Where can I find more info on the CELDT?**

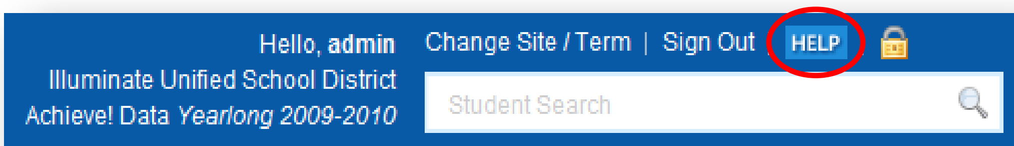
Visit http://www.cde.ca.gov/ta/tg/el/ for resources.

Where can I find more info on analyzing CELDT performance?

Visit the Help system's *Data Analysis* manual.

Where can I learn how to generate this report in my data system?

Visit the Help system's *Reports* manual.



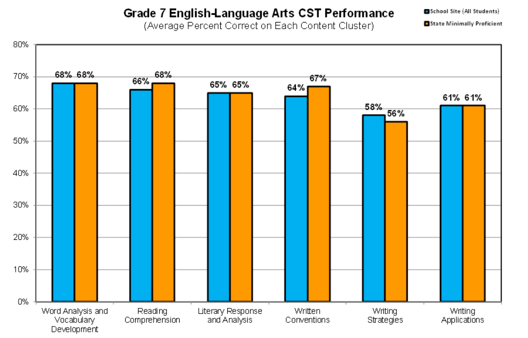
Who takes the CELDT and when?

All students whose home language is not English must test within 30 calendar days of enrolling in a California public school to determine classification as Fluent-English Proficient (FEP) or English Learner (EL). ELs must test every year thereafter until they are Reclassified as Fluent-English Proficient (R-FEP).

What do the performance levels mean?

1 = Beginning, 2 = Early Intermediate, 3 = Intermediate, 4 = Early Advanced, 5 = Advanced

REPORT 1



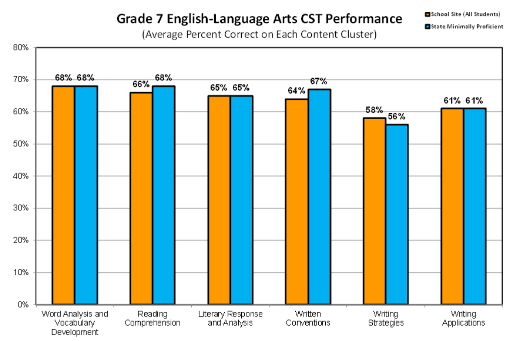
REPORT 2

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Student A	2	3	3	4	5	4
Student B	7	3	3	4	4	3
Student C	5	4	5	4	5	4
Student D	11	4	2	5	5	5
Average		3.5	3.3	4.3	4.8	4.0

Scenario 1: *Scenario 1 Participant (Control Group) Handouts*

REPORT 1



Warning: Clusters vary in difficulty, so the Site's highest % correct is not necessarily a strength.

What to Do: Site % – State Minimally Proficient % = # (highest # could be Site strength, lowest # could be Site weaknesses).

REPORT 2

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

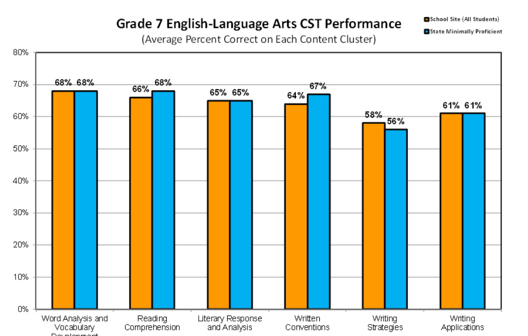
Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Student A	2	3	3	4	5	4
Student B	7	3	3	4	4	3
Student C	5	4	5	4	5	4
Student D	11	4	2	5	5	5
Average		3.5	3.3	4.3	4.8	4.0

Warning: "Overall" is not the only score that determines CELDT proficiency.

What to Do: Consider a student CELDT *Proficient* only with both:
✓ 4 or above *Overall*, &
✓ 3 or above in every domain

Scenario 2: *Scenario 2 (Footer A) Participant Handouts*

REPORT 1



Clusters vary in difficulty, so the Site's highest % correct is not necessarily a strength.
Compare the Site % to the State Minimally Proficient % (i.e., look at the degree to which the Site *beat* the SMP).
Site % - SMP % = # (cluster with highest difference could be Site strength, lowest difference could be Site weaknesses).

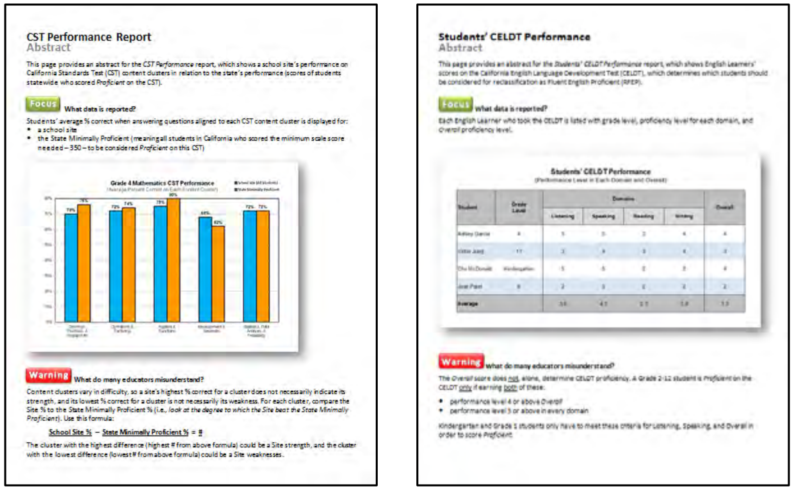
REPORT 2

Students' CELDT Performance
(Performance Level in Each Domain and Overall)

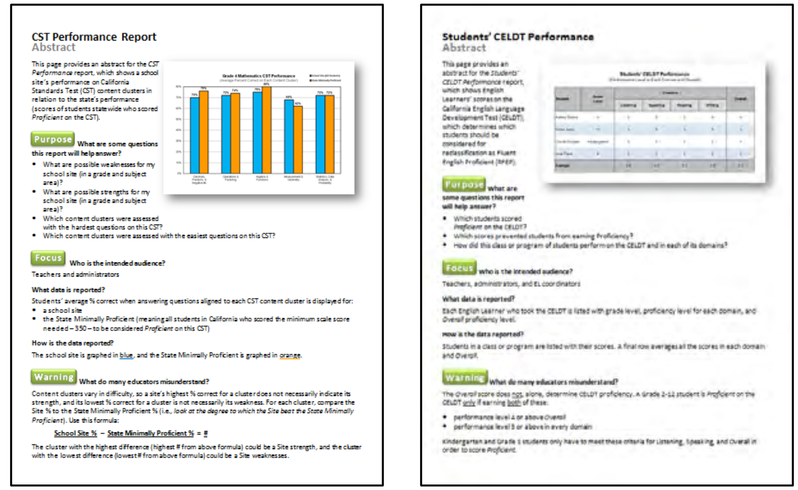
Student	Grade Level	Domains				Overall
		Listening	Speaking	Reading	Writing	
Student A	2	3	3	4	5	4
Student B	7	3	3	4	4	3
Student C	5	4	5	4	5	4
Student D	11	4	2	5	5	5
Average		3.5	3.3	4.3	4.8	4.0

The student's "Overall" score is not the only score that determines CELDT proficiency.
A student is *Proficient* on the CELDT only if earning both of these:
- performance level 4 or above *Overall*, &
- performance level 3 or above in every domain

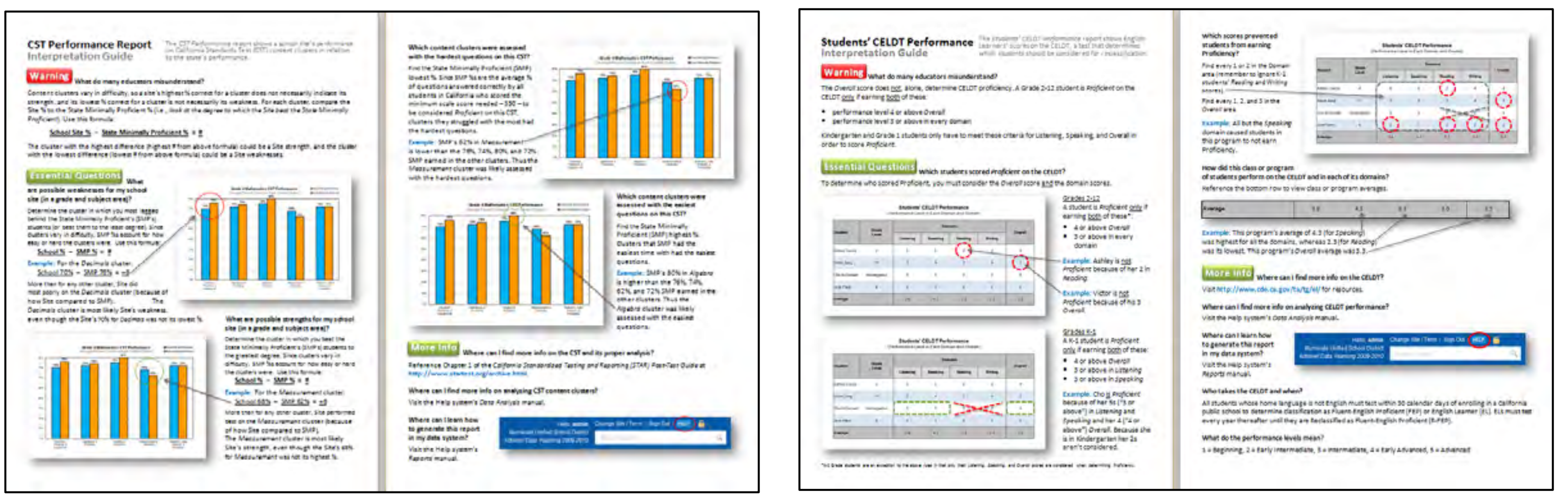
Scenario 3: *Scenario 3 (Footer B) Participant Handouts*



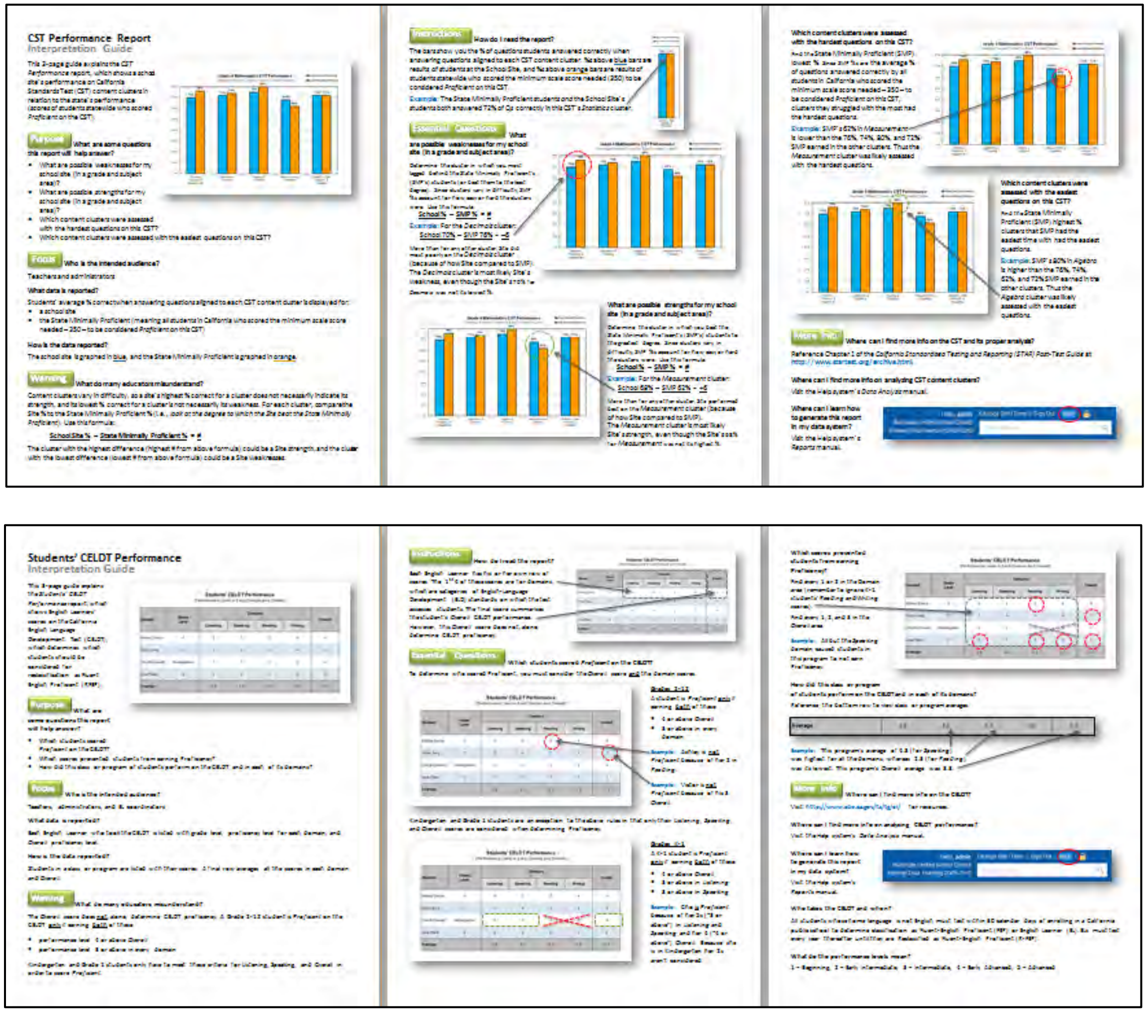
Scenario 4: *Scenario 4 Participant (Abstract A) Handouts; These Participants Also Received Scenario 1 Handouts*



Scenario 5: *Scenario 5 Participant (Abstract B) Handouts; These Participants Also Received Scenario 1 Handouts*



Scenario 6: *Scenario 6Participant (Interpretation Guide A) Handouts; These Participants Also Received Scenario 1 Handouts*



Scenario 7: *Scenario 7 Participant (Interpretation Guide B) Handouts; These Participants Also Received* Figure 1 *Handout*

Appendix D: Code Book for Respondent Data File

Column	Header/Label	Coding/Function	Respondent Row Contents
A	#	Added 1-211 from Earliest Row of Respondent Data (1) to Last Row of Respondent Data (211) to Record Original Order of Responses	Number
B	Timestamp	Automatically Added by Google Docs to Respondent's Original Data (Not Manipulated)	Date & Time
C	1. How long have you worked as an educator (e.g., teacher or administrator) for students under 19 years of age?	Respondent's Original Data (Not Manipulated)	Text
D	2. Which of the following roles best describes your current position?	Respondent's Original Data (Not Manipulated)	Text
E	3. How proficient are you at analyzing student performance data?	Respondent's Original Data (Not Manipulated)	Text
F	4. Which content cluster is most likely the School's strength?	Respondent's Original Data (Not Manipulated)	Text
G	5. Which content cluster is most likely the School's weakness?	Respondent's Original Data (Not Manipulated)	Text
H	6. Which student(s) did NOT score Proficient on the CELDT?	Respondent's Original Data (Not Manipulated)	Text

I	7. In which area(s) did at least 1 student earn a score that PREVENTED him/her from scoring Proficient on the CELDT?	Respondent's Original Data (Not Manipulated)	Text
J	What color is your folder?	Respondent's Original Data (Not Manipulated)	Text
K	8. The 2 reports you just used did not offer any special assistance in analyzing the data. If they had been accompanied by text (e.g., a footer, guide, or abstract) designed to help you interpret the data, would you likely have used the added support?	Applicable Respondent's Original Data (Not Manipulated)	Text
L	8. The 2 reports you just used contained footers with analysis guidelines designed to help you. Did you read these footers before answering questions related to the reports?	Applicable Respondent's Original Data (Not Manipulated)	Text
M	8. The 2 reports you just used were each accompanied by a 1-page abstract (like a reference sheet) with analysis guidelines designed to help you. Did you read these abstracts/sheets before answering questions related to the reports?	Applicable Respondent's Original Data (Not Manipulated)	Text
N	8. The 2 reports you just used were each accompanied by an interpretation guide (a packet) with analysis guidelines designed to help you. Did you read these guides before answering questions related to the reports?	Applicable Respondent's Original Data (Not Manipulated)	Text

O	9. Lots of professional development happens at school sites: for example, demonstrations to accompany textbook adoptions, meetings with colleagues to share differentiation strategies, training on how to use new software, etc. Only some professional development specifically focuses on how to analyze student data. Within the last 12 months, how many hours of professional development have you had that specifically focused on teaching you how to correctly interpret student data?	Respondent's Original Data (Not Manipulated)	Text
P	10. Educational Measurement refers to the analysis of student assessment data to draw conclusions about abilities. How many graduate-level courses have you taken that were specifically dedicated to educational measurement (e.g., student performance data analysis, measurement theory, or psychometrics)?	Respondent's Original Data (Not Manipulated)	Text
Q	Folder/Scenario	# Based on Same-Row Cell in Column J (Manually Added to Assist Coding: White=1, Green=2, Yellow=3, Purple=4, Blue=5, Black=6, Red=7)	Number
R	Support Use (Value)	Concatenated Values from Same-Row Cell in Columns K-O (Added to Assist Coding)	Text
S	School	Site Demographics (Manually Added After Each Administration)	Text

T	County	Site Demographics (Manually Added After Each Administration)	Text
U	City	Site Demographics (Manually Added After Each Administration)	Text
V	District	Site Demographics (Manually Added After Each Administration)	Text
W	2012 Growth API	Site Demographics (Manually Added After Each Administration)	Number
X	English Learners	Site Demographics (Manually Added After Each Administration)	Percentage
Y	Socioeconomically Disadvantaged	Site Demographics (Manually Added After Each Administration)	Percentage
Z	Students with Disabilities	Site Demographics (Manually Added After Each Administration)	Percentage
AA	1. How long have you worked as an educator (e.g., teacher or administrator) for students under 19 years of age?	Coded 1-5 Based on Same-Row Cell in Column C	Number
AB	2. Which of the following roles best describes your current position?	Coded 1-4 Based on Same-Row Cell in Column D	Number
AC	3. How proficient are you at analyzing student performance data?	Coded 1-4 Based on Same-Row Cell in Column E	Number

AD	4. Which content cluster is most likely the School's strength?	Coded 0-1 Based on Same-Row Cell in Column F	Number
AE	5. Which content cluster is most likely the School's weakness?	Coded 0-1 Based on Same-Row Cell in Column G	Number
AF	6. Which student(s) did NOT score Proficient on the CELDT?	Coded 0-1 Based on Same-Row Cell in Column H	Number
AG	7. In which area(s) did at least 1 student earn a score that PREVENTED him/her from scoring Proficient on the CELDT?	Coded 0-1 Based on Same-Row Cell in Column I	Number
AH	8. The 2 reports you just used did not offer any special assistance in analyzing the data. If they had been accompanied by text (e.g., a footer, guide, or abstract) designed to help you interpret the data, would you likely have used the added support?	Coded 1-2 Based on Same-Row Cell in Column K	Number
AI	8. The 2 reports you just used contained footers with analysis guidelines designed to help you. Did you read these footers before answering questions related to the reports?	Coded 1-4 Based on Same-Row Cell in Column L	Number
AJ	8. The 2 reports you just used were each accompanied by a 1-page abstract (like a reference sheet) with analysis guidelines designed to help you. Did you read these abstracts/sheets before answering questions related to the reports?	Coded 1-4 Based on Same-Row Cell in Column M	Number

AK	8. The 2 reports you just used were each accompanied by an interpretation guide (a packet) with analysis guidelines designed to help you. Did you read these guides before answering questions related to the reports?	Coded 1-4 Based on Same-Row Cell in Column N	Number
AL	Q8s Combined	Concatenated Values from Same-Row Cell in Columns AH-AK (Added to Assist Coding)	Number
AM	9. Lots of professional development happens at school sites: for example, demonstrations to accompany textbook adoptions, meetings with colleagues to share differentiation strategies, training on how to use new software, etc. Only some professional development specifically focuses on how to analyze student data. Within the last 12 months, how many hours of professional development have you had that specifically focused on teaching you how to correctly interpret student data?	Coded 1-5 Based on Same-Row Cell in Column O	Number
AN	10. Educational Measurement refers to the analysis of student assessment data to draw conclusions about abilities. How many graduate-level courses have you taken that were specifically dedicated to educational measurement (e.g., student performance data analysis, measurement theory, or psychometrics)?	Coded 1-5 Based on Same-Row Cell in Column P	Number

AO	677	% Correct by Site API (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AP	794	% Correct by Site API (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AQ	815	% Correct by Site API (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AR	827	% Correct by Site API (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AS	847	% Correct by Site API (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AT	891	% Correct by Site API (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AU	893	% Correct by Site API (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AV	895	% Correct by Site API (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage

AW	916	% Correct by Site API (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AX	8%	% Correct by Site % English Learner (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AY	10%	% Correct by Site % English Learner (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
AZ	16%	% Correct by Site % English Learner (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BA	27%	% Correct by Site % English Learner (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BB	30%	% Correct by Site % English Learner (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BC	33%	% Correct by Site % English Learner (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BD	38%	% Correct by Site % English Learner (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage

BE	45%	% Correct by Site % English Learner (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BF	46%	% Correct by Site % English Learner (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BG	22%	% Correct by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BH	23%	% Correct by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BI	31%	% Correct by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BJ	43%	% Correct by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BK	56%	% Correct by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BL	61%	% Correct by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage

BM	78%	% Correct by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BN	5%	% Correct by Site % Students with Disabilities (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BO	8%	% Correct by Site % Students with Disabilities (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BP	9%	% Correct by Site % Students with Disabilities (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BQ	10%	% Correct by Site % Students with Disabilities (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BR	11%	% Correct by Site % Students with Disabilities (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BS	12%	% Correct by Site % Students with Disabilities (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
BT	13%	% Correct by Site % Students with Disabilities (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage

BU	Buena Park Junior High	% Correct by Site Name (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion Name)	Percentage
BV	Charles G. Emery Elementary	% Correct by Site Name (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion Name)	Percentage
BW	Creek View Elementary	% Correct by Site Name (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion Name)	Percentage
BX	Etiwanda Colony Elementary	% Correct by Site Name (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion Name)	Percentage
BY	Grace Yokely Middle	% Correct by Site Name (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion Name)	Percentage
BZ	Hermosa Elementary	% Correct by Site Name (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion Name)	Percentage
CA	Ranch View Elementary	% Correct by Site Name (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion Name)	Percentage
CB	Rolling Ridge Elementary	% Correct by Site Name (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion Name)	Percentage

CC	Sylmar High	% Correct by Site Name (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion Name)	Percentage
CD	Elem	% Correct by Site School Level & School Level Type (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
CE	Mid/Jr	% Correct by Site School Level (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
CF	High	% Correct by Site School Level (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
CG	Secondary	% Correct by Site School Level Type (Added Value of Same-Row Cell in Column DK When Respondent Was from Site Matching Criterion)	Percentage
CH	< 1 yr	% Correct by Participant Veteran Status (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage
CI	At least 5 yrs	% Correct by Participant Veteran Status (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage

CJ	At least 10 yrs	% Correct by Participant Veteran Status (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage
CK	At least 15 yrs	% Correct by Participant Veteran Status (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage
CL	At least 20 yrs	% Correct by Participant Veteran Status (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage
CM	Teacher	% Correct by Participant Role (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column D Matched Criteria)	Percentage
CN	Colleague Coach	% Correct by Participant Role (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column D Matched Criteria)	Percentage
CO	Site Admin	% Correct by Participant Role (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column D Matched Criteria)	Percentage
CP	District Admin	% Correct by Participant Role (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column D Matched Criteria)	Percentage

CQ	Very Prof	% Correct by Participant Perceived Data Analysis Proficiency (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column E Matched Criteria)	Percentage
CR	Somewhat Prof	% Correct by Participant Perceived Data Analysis Proficiency (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column E Matched Criteria)	Percentage
CS	Not Prof	% Correct by Participant Perceived Data Analysis Proficiency (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column E Matched Criteria)	Percentage
CT	Far from Prof	% Correct by Participant Perceived Data Analysis Proficiency (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column E Matched Criteria)	Percentage
CU	0 hrs	% Correct by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage
CV	1 hr	% Correct by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage

CW	2 hrs	% Correct by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage
CX	5 hrs	% Correct by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage
CY	8 or more	% Correct by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage
CZ	0 courses	% Correct by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage
DA	1 course	% Correct by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage
DB	2 courses	% Correct by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage

DC	3 courses	% Correct by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage
DD	4 or more	% Correct by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column DK When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage
DE	% for #4	% Correct by Question (Coded 0% or 100% Based on 0 or 1 in Same-Row Cell in Column AD)	Percentage
DF	% for #5	% Correct by Question (Coded 0% or 100% Based on 0 or 1 in Same-Row Cell in Column AE)	Percentage
DG	% for Rpt 1	% Correct by Report (Mean/Averaged Values from Same-Row Cell in Columns DE-DF)	Percentage
DH	% for #6	% Correct by Question (Coded 0% or 100% Based on 0 or 1 in Same-Row Cell in Column AF)	Percentage
DI	% for #7	% Correct by Question (Coded 0% or 100% Based on 0 or 1 in Same-Row Cell in Column AG)	Percentage
DJ	% for Rpt 2	% Correct by Report (Mean/Averaged Values from Same-Row Cell in Columns DH-DI)	Percentage
DK	% Overall	% Correct Overall (Mean/Averaged Values from Same-Row Cell in Columns DE, DF, DH, and DI)	Percentage

DL	No Sup %	% Correct by Scenario (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 1)	Percentage
DM	Support %	% Correct by Scenario (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 2-7)	Empty (Contents Only in Summary Rows)
DN	Scenario 1 %	% Correct by Scenario (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 1)	Percentage
DO	Scenario 2 %	% Correct by Scenario (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 2)	Percentage
DP	Scenario 3 %	% Correct by Scenario (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 1)	Percentage
DQ	Scenario 2+3 % (Footer)	% Correct by Scenario (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DO-DP)	Empty (Contents Only in Summary Rows)
DR	Scenario 4 %	% Correct by Scenario (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 4)	Percentage

DS	Scenario 5 %	% Correct by Scenario (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 5)	Percentage
DT	Scenario 4+5 % (Abstract)	% Correct by Scenario (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DR-DS)	Empty (Contents Only in Summary Rows)
DU	Scenario 6 %	% Correct by Scenario (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 6)	Percentage
DV	Scenario 7 %	% Correct by Scenario (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 7)	Percentage
DW	Scenario 6+7 % (Guide)	% Correct by Scenario (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DU-DV)	Empty (Contents Only in Summary Rows)
DX	Control: Wouldn't Use %	% Correct by Control Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 1 and Same-Row Cell in Column AH Matched 1)	Percentage

DY	Control: Would Use %	% Correct by Control Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 1 and Same-Row Cell in Column AH Matched 2)	Percentage
DZ	Footer A: No Use, Q4-5 R1	% Correct by Footer A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 2 and Same-Row Cell in Column AI Matched 1 or 3)	Percentage
EA	Footer A: No Use, Q6-7 R2	% Correct by Footer A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 2 and Same-Row Cell in Column AI Matched 1 or 2)	Percentage
EB	Footer A: No Use	% Correct by Footer A Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DZ-EA)	Empty (Contents Only in Summary Rows)
EC	Footer A: Use Q4-5 R1	% Correct by Footer A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 2 and Same-Row Cell in Column AI Matched 2 or 4)	Percentage
ED	Footer A: Use Q6-7 R2	% Correct by Footer A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 2 and Same-Row Cell in Column AI Matched 3 or 4)	Percentage

EE	Footer A: Use	% Correct by Footer A Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EC-ED)	Empty (Contents Only in Summary Rows)
EF	Footer B: No Use, Q4-5 R1	% Correct by Footer B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 3 and Same-Row Cell in Column AI Matched 1 or 3)	Percentage
EG	Footer B: No Use, Q6-7 R2	% Correct by Footer B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 3 and Same-Row Cell in Column AI Matched 1 or 2)	Percentage
EH	Footer B: No Use	% Correct by Footer B Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DZ-EA)	Empty (Contents Only in Summary Rows)
EI	Footer B: Use Q4-5 R1	% Correct by Footer B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 3 and Same-Row Cell in Column AI Matched 2 or 4)	Percentage
EJ	Footer B: Use Q6-7 R2	% Correct by Footer B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 3 and Same-Row Cell in Column AI Matched 3 or 4)	Percentage

EK	Footer B: Use	% Correct by Footer B Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EC-ED)	Empty (Contents Only in Summary Rows)
EL	No Use	% Correct by Footer Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DZ, EA, EF, & EG)	Empty (Contents Only in Summary Rows)
EM	Use	% Correct by Footer Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EC, ED, EI, & EJ)	Empty (Contents Only in Summary Rows)
EN	Abstract A: No Use, Q4-5 R1	% Correct by Abstract A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 4 and Same-Row Cell in Column AJ Matched 1 or 3)	Percentage
EO	Abstract A: No Use, Q6-7 R2	% Correct by Abstract A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 4 and Same-Row Cell in Column AJ Matched 1 or 2)	Percentage

EP	Abstract A: No Use	% Correct by Abstract A Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DZ-EA)	Empty (Contents Only in Summary Rows)
EQ	Abstract A: Use Q4-5 R1	% Correct by Abstract A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 4 and Same-Row Cell in Column AJ Matched 2 or 4)	Percentage
ER	Abstract A: Use Q6-7 R2	% Correct by Abstract A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 4 and Same-Row Cell in Column AJ Matched 3 or 4)	Percentage
ES	Abstract A: Use	% Correct by Abstract A Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EC-ED)	Empty (Contents Only in Summary Rows)
ET	Abstract B: No Use, Q4-5 R1	% Correct by Abstract B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 5 and Same-Row Cell in Column AJ Matched 1 or 3)	Percentage
EU	Abstract B: No Use, Q6-7 R2	% Correct by Abstract B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 5 and Same-Row Cell in Column AJ Matched 1 or 2)	Percentage

EV	Abstract B: No Use	% Correct by Abstract B Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DZ-EA)	Empty (Contents Only in Summary Rows)
EW	Abstract B: Use Q4-5 R1	% Correct by Abstract B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 5 and Same-Row Cell in Column AJ Matched 2 or 4)	Percentage
EX	Abstract B: Use Q6-7 R2	% Correct by Abstract B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 5 and Same-Row Cell in Column AJ Matched 3 or 4)	Percentage
EY	Abstract B: Use	% Correct by Abstract B Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EC-ED)	Empty (Contents Only in Summary Rows)
EZ	No Use	% Correct by Abstract Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EN, EO, ET, & EU)	Empty (Contents Only in Summary Rows)

FA	Use	% Correct by Abstract Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EQ, ER, EW, & EX)	Empty (Contents Only in Summary Rows)
FB	Interpretation Guide A: No Use, Q4-5 R1	% Correct by Interpretation Guide A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 6 and Same-Row Cell in Column AK Matched 1 or 3)	Percentage
FC	Interpretation Guide A: No Use, Q6-7 R2	% Correct by Interpretation Guide A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 6 and Same-Row Cell in Column AK Matched 1 or 2)	Percentage
FD	Interpretation Guide A: No Use	% Correct by Interpretation Guide A Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DZ-EA)	Empty (Contents Only in Summary Rows)
FE	Interpretation Guide A: Use Q4-5 R1	% Correct by Interpretation Guide A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 6 and Same-Row Cell in Column AK Matched 2 or 4)	Percentage
FF	Interpretation Guide A: Use Q6-7 R2	% Correct by Interpretation Guide A Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 6 and Same-Row Cell in Column AK Matched 3 or 4)	Percentage

FG	Interpretation Guide A: Use	% Correct by Interpretation Guide A Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EC-ED)	Empty (Contents Only in Summary Rows)
FH	Interpretation Guide B: No Use, Q4-5 R1	% Correct by Interpretation Guide B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 7 and Same-Row Cell in Column AK Matched 1 or 3)	Percentage
FI	Interpretation Guide B: No Use, Q6-7 R2	% Correct by Interpretation Guide B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 7 and Same-Row Cell in Column AK Matched 1 or 2)	Percentage
FJ	Interpretation Guide B: No Use	% Correct by Interpretation Guide B Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DZ-EA)	Empty (Contents Only in Summary Rows)
FK	Interpretation Guide B: Use Q4-5 R1	% Correct by Interpretation Guide B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 7 and Same-Row Cell in Column AK Matched 2 or 4)	Percentage
FL	Interpretation Guide B: Use Q6-7 R2	% Correct by Interpretation Guide B Use (Added Value of Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 7 and Same-Row Cell in Column AK Matched 3 or 4)	Percentage

FM	Interpretation Guide B: Use	% Correct by Interpretation Guide B Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EC-ED)	Empty (Contents Only in Summary Rows)
FN	No Use	% Correct by Interpretation Guide Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns FB, FC, FH, & FI)	Empty (Contents Only in Summary Rows)
FO	Use	% Correct by Interpretation Guide Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns FE, FF, FK, & FL)	Empty (Contents Only in Summary Rows)
FP	Supports Not Used	% Correct by Support Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DL, DZ, EA, EF, EG, EN, EO, ET, EU, FB, FC, FH, & FI)	Empty (Contents Only in Summary Rows)
FQ	Supports Avail. But Not Used	% Correct by Support Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DZ, EA, EF, EG, EN, EO, ET, EU, FB, FC, FH, & FI)	Empty (Contents Only in Summary Rows)

FR	Supports Used	% Correct by Support Use (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns EC, ED, EI, EJ, EQ, ER, EW, EX, FE, FF, FK, & FL)	Empty (Contents Only in Summary Rows)
FS	677	Instance/Likelihood of Using Support by Site API (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
FT	794	Instance/Likelihood of Using Support by Site API (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
FU	815	Instance/Likelihood of Using Support by Site API (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
FV	827	Instance/Likelihood of Using Support by Site API (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
FW	847	Instance/Likelihood of Using Support by Site API (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
FX	891	Instance/Likelihood of Using Support by Site API (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage

FY	893	Instance/Likelihood of Using Support by Site API (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
FZ	895	Instance/Likelihood of Using Support by Site API (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GA	916	Instance/Likelihood of Using Support by Site API (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GB	8%	Instance/Likelihood of Using Support by Site % English Learner (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GC	10%	Instance/Likelihood of Using Support by Site % English Learner (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GD	16%	Instance/Likelihood of Using Support by Site % English Learner (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GE	27%	Instance/Likelihood of Using Support by Site % English Learner (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GF	30%	Instance/Likelihood of Using Support by Site % English Learner (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage

GG	33%		Instance/Likelihood of Using Support by Site % English Learner (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GH	38%		Instance/Likelihood of Using Support by Site % English Learner (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GI	45%		Instance/Likelihood of Using Support by Site % English Learner (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GJ	46%		Instance/Likelihood of Using Support by Site % English Learner (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GK	22%		Instance/Likelihood of Using Support by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GL	23%		Instance/Likelihood of Using Support by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GM	31%		Instance/Likelihood of Using Support by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage

GN	43%	Instance/Likelihood of Using Support by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GO	56%	Instance/Likelihood of Using Support by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GP	61%	Instance/Likelihood of Using Support by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GQ	78%	Instance/Likelihood of Using Support by Site % Socioeconomically Disadvantaged (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GR	5%	Instance/Likelihood of Using Support by Site % Students with Disabilities (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GS	8%	Instance/Likelihood of Using Support by Site % Students with Disabilities (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage

GT	9%	Instance/Likelihood of Using Support by Site % Students with Disabilities (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GU	10%	Instance/Likelihood of Using Support by Site % Students with Disabilities (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GV	11%	Instance/Likelihood of Using Support by Site % Students with Disabilities (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GW	12%	Instance/Likelihood of Using Support by Site % Students with Disabilities (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GX	13%	Instance/Likelihood of Using Support by Site % Students with Disabilities (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
GY	Elem	Instance/Likelihood of Using Support by Site School Level & School Level Type (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage

GZ	Mid/Jr	Instance/Likelihood of Using Support by Site School Level (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
HA	High	Instance/Likelihood of Using Support by Site School Level (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
HB	Secondary	Instance/Likelihood of Using Support by Site School Level Type (Added Value of Same-Row Cell in Column IO When Respondent Was from Site Matching Criterion)	Percentage
HC	< 1 yr	Instance/Likelihood of Using Support by Participant Veteran Status (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage
HD	At least 5 yrs	Instance/Likelihood of Using Support by Participant Veteran Status (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage
HE	At least 10 yrs	Instance/Likelihood of Using Support by Participant Veteran Status (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage
HF	At least 15 yrs	Instance/Likelihood of Using Support by Participant Veteran Status (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage

HG	At least 20 yrs	Instance/Likelihood of Using Support by Participant Veteran Status (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column C Matched Criteria)	Percentage
HH	Teacher	Instance/Likelihood of Using Support by Participant Role (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column D Matched Criteria)	Percentage
HI	Colleague Coach	Instance/Likelihood of Using Support by Participant Role (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column D Matched Criteria)	Percentage
HJ	Site Admin	Instance/Likelihood of Using Support by Participant Role (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column D Matched Criteria)	Percentage
HK	District Admin	Instance/Likelihood of Using Support by Participant Role (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column D Matched Criteria)	Percentage
HL	Very Prof	Instance/Likelihood of Using Support by Participant Perceived Data Analysis Proficiency (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column E Matched Criteria)	Percentage

HM	Somewhat Prof	Instance/Likelihood of Using Support by Participant Perceived Data Analysis Proficiency (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column E Matched Criteria)	Percentage
HN	Not Prof	Instance/Likelihood of Using Support by Participant Perceived Data Analysis Proficiency (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column E Matched Criteria)	Percentage
HO	Far from Prof	Instance/Likelihood of Using Support by Participant Perceived Data Analysis Proficiency (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column E Matched Criteria)	Percentage
HP	0 hrs	Instance/Likelihood of Using Support by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage
HQ	1 hr	Instance/Likelihood of Using Support by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage

HR	2 hrs	Instance/Likelihood of Using Support by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage
HS	5 hrs	Instance/Likelihood of Using Support by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage
HT	8 or more	Instance/Likelihood of Using Support by Participant PD in Data Analysis (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column O Matched Criteria)	Percentage
HU	0 courses	Instance/Likelihood of Using Support by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage
HV	1 course	Instance/Likelihood of Using Support by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage

HW	2 courses	Instance/Likelihood of Using Support by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage
HX	3 courses	Instance/Likelihood of Using Support by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage
HY	4 or more	Instance/Likelihood of Using Support by Participant Graduate Courses in Educational Measurement (Added Value of Same-Row Cell in Column IO When Respondent Answer in Same-Row Cell in Column P Matched Criteria)	Percentage
HZ	No Sup %	Instance/Likelihood of Using Support by Scenario (Added Value of Same-Row Cell in Column IO When Same-Row Cell in Column Q Matched 1)	Percentage
IA	Support %	Instance/Likelihood of Using Support by Scenario (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Column DK When Same-Row Cell in Column Q Matched 2-7)	Percentage
IB	Scenario 1 %	Instance/Likelihood of Using Support by Scenario (Added Value of Same-Row Cell in Column IO When Same-Row Cell in Column Q Matched 1)	Percentage

IC	Scenario 2 %	Instance/Likelihood of Using Support by Scenario (Added Value of Same-Row Cell in Column IO When Same-Row Cell in Column Q Matched 2)	Percentage
ID	Scenario 3 %	Instance/Likelihood of Using Support by Scenario (Added Value of Same-Row Cell in Column IO When Same-Row Cell in Column Q Matched 1)	Percentage
IE	Scenario 2+3 % (Footer)	Instance/Likelihood of Using Support by Scenario (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DO-DP)	Percentage
IF	Scenario 4 %	Instance/Likelihood of Using Support by Scenario (Added Value of Same-Row Cell in Column IO When Same-Row Cell in Column Q Matched 4)	Percentage
IG	Scenario 5 %	Instance/Likelihood of Using Support by Scenario (Added Value of Same-Row Cell in Column IO When Same-Row Cell in Column Q Matched 5)	Percentage
IH	Scenario 4+5 % (Abstract)	Instance/Likelihood of Using Support by Scenario (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DR-DS)	Percentage
II	Scenario 6 %	Instance/Likelihood of Using Support by Scenario (Added Value of Same-Row Cell in Column IO When Same-Row Cell in Column Q Matched 6)	Percentage
IJ	Scenario 7 %	Instance/Likelihood of Using Support by Scenario (Added Value of Same-Row Cell in Column IO When Same-Row Cell in Column Q Matched 7)	Percentage

IK	Scenario 6+7 % (Guide)	Instance/Likelihood of Using Support by Scenario (No Data, Except in Summary Row at Bottom Averaging Every Value in Same-Row Cell in Columns DU-DV)	Percentage
IL	Support Use (Code)	Coded 1-4 Based on Same-Row Cell in Columns Q and AL	Number
IM	R1 Support Use (%)	Instance/Likelihood of Using Support by Report (Added 0% or 100% Based on Whether Value of Same-Row Cell in Column AL Indicated Support Was Used for Report 1)	Percentage
IN	R2 Support Use (%)	Instance/Likelihood of Using Support by Report (Added 0% or 100% Based on Whether Value of Same-Row Cell in Column AL Indicated Support Was Used for Report 2)	Percentage
IO	Support Use (%)	Instance/Likelihood of Using Support (Added % to Match Value of Same-Row Cell in Column AL)	Percentage
IP	Support Access	Whether or Not Support Was Present (Added 0 When Same-Row Cell in Column Q Matched 1 and Added 1 When Same-Row Cell in Column Q Matched 2-7)	Number
IQ	For Use: Q4 Correct?	Coded 0-1 Based on Same-Row Cell Value in Column DE (100% Becomes 1, 0% Becomes 0)	Number
IR	For Use: Q4 Support Used?	Coded 0 (Not Used) or 1 (Used) Based on Value of Same-Row Cell in Column AL and Its Indication of Wether or Not Support Was Used for Report 1 Questions (1 and 2 Column AL Values Become 0, 3 and 4 Column AL Values Become 1, and all Rows with 1 in Column Q Become 0)	Number

IS	For Use: Q5 Correct?	Coded 0-1 Based on Same-Row Cell Value in Column DF (100% Becomes 1, 0% Becomes 0)	Number
IT	For Use: Q5 Support Used?	Added Value of Same-Row Cell in Column IR (to Minimize Human Error When Pasting Paired Data from Columns IS and IT into PASW SPSS)	Number
IU	For Use: Q6 Correct?	Coded 0-1 Based on Same-Row Cell Value in Column DH (100% Becomes 1, 0% Becomes 0)	Number
IV	For Use: Q6 Support Used?	Coded 0 (Not Used) or 1 (Used) Based on Value of Same-Row Cell in Column AL and Its Indication of Wether or Not Support Was Used for Report 2 Questions (1 and 3 Column AL Values Become 0, 2 and 4 Column AL Values Become 1, and all Rows with 1 in Column Q Become 0)	Number
IW	For Use: Q7 Correct?	Coded 0-1 Based on Same-Row Cell Value in Column DI (100% Becomes 1, 0% Becomes 0)	Number
IX	For Use: Q7 Support Used?	Added Value of Same-Row Cell in Column IV (to Minimize Human Error When Pasting Paired Data from Columns IS and IT into PASW SPSS)	Number
IY	For Access: Q4 Correct?	Added Value of Same-Row Cell in Column IQ (to Minimize Human Error When Pasting Paired Data from Columns IY and IZ into PASW SPSS)	Number
IZ	For Access: Q4 Support Access?	Added Value of Same-Row Cell in Column IP (to Minimize Human Error When Pasting Paired Data from Columns IY and IZ into PASW SPSS)	Number

JA	For Access: Q5 Correct?	Added Value of Same-Row Cell in Column IS (to Minimize Human Error When Pasting Paired Data from Columns JA and JB into PASW SPSS)	Number
JB	For Access: Q5 Support Access?	Added Value of Same-Row Cell in Column IP (to Minimize Human Error When Pasting Paired Data from Columns JA and JB into PASW SPSS)	Number
JC	For Access: Q6 Correct?	Added Value of Same-Row Cell in Column IU (to Minimize Human Error When Pasting Paired Data from Columns JC and JD into PASW SPSS)	Number
JD	For Access: Q6 Support Access?	Added Value of Same-Row Cell in Column IP (to Minimize Human Error When Pasting Paired Data from Columns JC and JD into PASW SPSS)	Number
JE	For Access: Q7 Correct?	Added Value of Same-Row Cell in Column IW (to Minimize Human Error When Pasting Paired Data from Columns JE and JF into PASW SPSS)	Number
JF	For Access: Q7 Support Access?	Added Value of Same-Row Cell in Column IP (to Minimize Human Error When Pasting Paired Data from Columns JE and JF into PASW SPSS)	Number
JG	School Level Type	Coded 1-2 (Elementary-Secondary) Based on Value of Same-Row Cell in Column S	Number
JH	School Level	Coded 1-3 (Elementary-High) Based on Value of Same-Row Cell in Column S	Number

Appendix E: Independent Samples T-Test for Support Use

Group Statistics

	Support Use (0 Not Used, 1 Used)	N	Mean
Analysis Accuracy (% Correct)	0	426	.07
	1	418	.45

Group Statistics

	Support Use (0 Not Used, 1 Used)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	0	.260	.013
	1	.499	.024

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	1059.423	.000
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	-13.985	842	.000
	Equal variances not assumed	-13.910	625.660	.000

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	-.382	.027
	Equal variances not assumed	-.382	.027

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-.435	-.328
	Equal variances not assumed	-.436	-.328

Appendix F: Independent Samples T-Test for Footer Use

Group Statistics

	Footer Use (Not Used, 1 Used)	N	Mean
Analysis Accuracy (% Correct)	0	190	.09
	1	174	.44

Group Statistics

	Footer Use (Not Used, 1 Used)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	0	.294	.021
	1	.498	.038

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	302.184	.000
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	-8.195	362	.000
	Equal variances not assumed	-8.022	275.119	.000

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	-.348	.042
	Equal variances not assumed	-.348	.043

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-.431	-.264
	Equal variances not assumed	-.433	-.262

Appendix G: Independent Samples T-Test for Abstract Use

Group Statistics

	Abstract Use (0 Not Used, 1 Used)	N	Mean
Analysis Accuracy (% Correct)	0	244	.10
	1	120	.37

Group Statistics

	Abstract Use (0 Not Used, 1 Used)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	0	.298	.019
	1	.484	.044

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	150.505	.000
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	-6.507	362	.000
	Equal variances not assumed	-5.575	164.850	.000

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	-.268	.041
	Equal variances not assumed	-.268	.048

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-.349	-.187
	Equal variances not assumed	-.363	-.173

Appendix H: Independent Samples T-Test for Interpretation Guide Use

Group Statistics

	Interp. Guide Use (0 Not Used, 1 Used)	N	Mean
Analysis Accuracy (% Correct)	0	240	.07
	1	124	.56

Group Statistics

	Interp. Guide Use (0 Not Used, 1 Used)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	0	.257	.017
	1	.499	.045

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	322.455	.000
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	-12.265	362	.000
	Equal variances not assumed	-10.166	157.550	.000

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	-.486	.040
	Equal variances not assumed	-.486	.048

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-.563	-.408
	Equal variances not assumed	-.580	-.391

Appendix I: Independent Samples T-Test for Support Presence

Group Statistics

	Support Presence (0 Not Present, 1 Present)	N	Mean
Analysis Accuracy (% Correct)	0	124	.11
	1	720	.29

Group Statistics

	Support Presence (0 Not Present, 1 Present)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	0	.318	.029
	1	.453	.017

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	114.558	.000
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	-4.121	842	.000
	Equal variances not assumed	-5.266	219.531	.000

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	-.175	.042
	Equal variances not assumed	-.175	.033

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-.258	-.091
	Equal variances not assumed	-.240	-.109

Appendix J: Independent Samples T-Test for Footer Presence

Group Statistics

	Footer Presence (0 Not Present, 1 Present)	N	Mean
Analysis Accuracy (% Correct)	0	124	.11
	1	240	.34

Group Statistics

	Footer Presence (0 Not Present, 1 Present)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	0	.318	.029
	1	.474	.031

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	137.571	.000
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	-4.753	362	.000
	Equal variances not assumed	-5.369	338.226	.000

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	-.225	.047
	Equal variances not assumed	-.225	.042

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-.318	-.132
	Equal variances not assumed	-.307	-.142

Appendix K: Independent Samples T-Test for Abstract Presence

Group Statistics

	Abstract Presence (0 Not Present, 1 Present)	N	Mean
Analysis Accuracy (% Correct)	0	124	.11
	1	240	.23

Group Statistics

	Abstract Presence (0 Not Present, 1 Present)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	0	.318	.029
	1	.418	.027

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	32.438	.000
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	-2.618	362	.009
	Equal variances not assumed	-2.853	312.890	.005

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	-.112	.043
	Equal variances not assumed	-.112	.039

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-.196	-.028
	Equal variances not assumed	-.189	-.035

Appendix L: Independent Samples T-Test for Interpretation Guide Presence

Group Statistics

	Interp. Guide Presence (0 Not Present, 1 Present)	N	Mean
Analysis Accuracy (% Correct)	0	124	.11
	1	240	.30

Group Statistics

	Interp. Guide Presence (0 Not Present, 1 Present)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	0	.318	.029
	1	.459	.030

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	92.109	.000
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	-4.061	362	.000
	Equal variances not assumed	-4.547	332.451	.000

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	-.187	.046
	Equal variances not assumed	-.187	.041

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-.278	-.096
	Equal variances not assumed	-.268	-.106

Appendix M: Independent Samples T-Test for Footer Format

Group Statistics

	Footer Format (2 Shorter, 3 Longer)	N	Mean
Analysis Accuracy (% Correct)	2	30	35.83
	3	30	31.67

Group Statistics

	Footer Format (2 Shorter, 3 Longer)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	2	32.618	5.955
	3	33.434	6.104

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	.063	.803
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	.489	58	.627
	Equal variances not assumed	.489	57.965	.627

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	4.167	8.528
	Equal variances not assumed	4.167	8.528

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-12.904	21.237
	Equal variances not assumed	-12.904	21.237

Appendix N: Independent Samples T-Test for Abstract Format

Group Statistics

	Abstract Format (4 Less Dense, 5 Denser)	N	Mean	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	4	30	20.83	27.919	5.097
	5	30	24.17	36.248	6.618

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	.832	.365
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	-.399	58	.691
	Equal variances not assumed	-.399	54.450	.691

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	-3.333	8.353
	Equal variances not assumed	-3.333	8.353

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-20.055	13.388
	Equal variances not assumed	-20.078	13.411

Appendix O: Independent Samples T-Test for Interpretation Guide Format

Group Statistics

	Interp. Guide Format (6 2-Page, 7 3-Page)	N	Mean
Analysis Accuracy (% Correct)	6	30	31.67
	7	30	28.33

Group Statistics

	Interp. Guide Format (6 2-Page, 7 3-Page)	Std. Deviation	Std. Error Mean
Analysis Accuracy (% Correct)	6	37.677	6.879
	7	29.165	5.325

Independent Samples Test

		Levene's Test for Equality of Variances	
		F	Sig.
Analysis Accuracy (% Correct)	Equal variances assumed	2.165	.147
	Equal variances not assumed		

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Analysis Accuracy (% Correct)	Equal variances assumed	.383	58	.703
	Equal variances not assumed	.383	54.572	.703

Independent Samples Test

		t-test for Equality of Means	
		Mean Difference	Std. Error Difference
Analysis Accuracy (% Correct)	Equal variances assumed	3.333	8.699
	Equal variances not assumed	3.333	8.699

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
Analysis Accuracy (% Correct)	Equal variances assumed	-14.079	20.746
	Equal variances not assumed	-14.103	20.769

Appendix P: Crosstabulated Chi-Square Tests for Variable Relationship with Data

Analysis Accuracy

School Level Type

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
School Level Type (1 Elem., 2 Sec.) * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

School Level Type (1 Elem., 2 Sec.) * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
School Level Type (1 Elem., 2 Sec.)	1	67	10	16	37	2	132
	2	42	7	7	19	4	79
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3.122[a]	4	.538
Likelihood Ratio	3.048	4	.550
N of Valid Cases	211		

a. 2 cells (20.0%) have expected count less than 5. The minimum expected count is 2.25.

School Level

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
School Level (1 Elem., 2 Mid./Jr., 3 High) * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

School Level (1 Elem., 2 Mid./Jr., 3 High) * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
School Level (1 Elem., 2 Mid./Jr., 3 High)	1	67	10	16	37	2	132
	2	26	4	4	12	1	47
	3	16	3	3	7	3	32
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.869[a]	8	.551
Likelihood Ratio	5.251	8	.730
N of Valid Cases	211		

a. 6 cells (40.0%) have expected count less than 5. The minimum expected count is .91.

Academic Performance

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
API * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

API * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
API	677	16	3	3	7	3	32
	794	20	1	4	8	0	33
	815	13	1	0	10	0	24
	827	6	3	0	4	1	14
	847	11	1	3	7	0	22
	891	14	3	3	8	0	28
	893	8	2	2	4	0	16
	895	13	3	5	8	2	31
	916	8	0	3	0	0	11
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	33.439[a]	32	.397
Likelihood Ratio	40.837	32	.136
N of Valid Cases	211		

a. 30 cells (66.7%) have expected count less than 5. The minimum expected count is .31.

English Learner Population

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
English Learner * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

English Learner * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
English Learner	10%	13	3	5	8	2	31
	16%	8	0	3	0	0	11
	27%	11	1	3	7	0	22
	30%	20	1	4	8	0	33
	33%	13	1	0	10	0	24
	38%	16	3	3	7	3	32
	45%	6	3	0	4	1	14
	46%	14	3	3	8	0	28
	8%	8	2	2	4	0	16
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	33.439[a]	32	.397
Likelihood Ratio	40.837	32	.136
N of Valid Cases	211		

a. 30 cells (66.7%) have expected count less than 5. The minimum expected count is .31.

Socioeconomically Disadvantaged Population

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Socioeconomically Disadvantaged * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

Socioeconomically Disadvantaged * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
Socioeconomically Disadvantaged	22%	8	0	3	0	0	11
	23%	13	3	5	8	2	31
	31%	8	2	2	4	0	16
	43%	14	3	3	8	0	28
	56%	11	1	3	7	0	22
	61%	33	2	4	18	0	57
	78%	22	6	3	11	4	46
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	26.870[a]	24	.311
Likelihood Ratio	31.484	24	.140
N of Valid Cases	211		

a. 21 cells (60.0%) have expected count less than 5. The minimum expected count is .31.

Students with Disabilities Population

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Students with Disabilities * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

Students with Disabilities * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
Students with Disabilities	10%	19	1	6	7	0	33
	11%	20	1	4	8	0	33
	12%	16	3	3	7	3	32
	13%	13	3	5	8	2	31
	5%	8	2	2	4	0	16
	8%	14	3	3	8	0	28
	9%	19	4	0	14	1	38
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	22.823[a]	24	.530
Likelihood Ratio	27.941	24	.263
N of Valid Cases	211		

a. 22 cells (62.9%) have expected count less than 5. The minimum expected count is .45.

Veteran Status

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Veteran Status * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

Veteran Status * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
Veteran Status	10 years	15	5	6	5	2	33
	15 years	32	6	7	22	0	67
	20 or more years	53	4	9	20	3	89
	5 years	8	2	1	8	1	20
	less than 1 year	1	0	0	1	0	2
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16.879[a]	16	.393
Likelihood Ratio	18.578	16	.291
N of Valid Cases	211		

a. 13 cells (52.0%) have expected count less than 5. The minimum expected count is .06.

Role

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Role * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

Role * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
Role	Colleague Coach (e.g., Teacher on Special Assignment)	1	0	0	1	0	2
	District Administrator	0	1	0	1	0	2
	Site/School Administrator	5	1	2	0	0	8
	Teacher	103	15	21	54	6	199
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11.266[a]	12	.506
Likelihood Ratio	12.360	12	.417
N of Valid Cases	211		

a. 15 cells (75.0%) have expected count less than 5. The minimum expected count is .06.

Perceived Data Analysis Accuracy

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Perceived Data Analysis Proficiency * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

Perceived Data Analysis Proficiency * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)			
		0%	100%	25%	50%
Perceived Data Analysis Proficiency	Far from proficient	4	0	0	1
	Not proficient	12	1	3	5
	Somewhat proficient	70	13	17	36
	Very proficient	23	3	3	14
Total		109	17	23	56

Perceived Data Analysis Proficiency * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)	
		75%	Total
Perceived Data Analysis Proficiency	Far from proficient	0	5
	Not proficient	1	22
	Somewhat proficient	3	139
	Very proficient	2	45
Total		6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.238[a]	12	.950
Likelihood Ratio	6.293	12	.901
N of Valid Cases	211		

a. 12 cells (60.0%) have expected count less than 5. The minimum expected count is .14.

Professional Development (PD)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
PD * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

PD * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
PD	0 hours	50	6	7	22	2	87
	1 hour	24	3	6	13	2	48
	2 hours	16	3	7	11	2	39
	5 hours	11	1	1	6	0	19
	8 pr more	8	4	2	4	0	18
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.441[a]	16	.713
Likelihood Ratio	11.853	16	.754
N of Valid Cases	211		

a. 13 cells (52.0%) have expected count less than 5. The minimum expected count is .51.

Graduate Educational Measurement Courses

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Courses * Analysis Accuracy (% Correct)	211	100.0%	0	.0%	211	100.0%

Courses * Analysis Accuracy (% Correct) Crosstabulation

Count

		Analysis Accuracy (% Correct)					Total
		0%	100%	25%	50%	75%	
Courses	0 courses	55	7	13	24	1	100
	1 course	22	4	7	15	3	51
	2 courses	19	5	2	8	1	35
	3 courses	6	0	0	4	1	11
	4 or more	7	1	1	5	0	14
Total		109	17	23	56	6	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.938[a]	16	.677
Likelihood Ratio	14.678	16	.548
N of Valid Cases	211		

a. 14 cells (56.0%) have expected count less than 5. The minimum expected count is .31.

Appendix Q: Crosstabulated Chi-Square Tests for Variable Relationship with Support Use

School Level Type

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
School Level Type (1 Elem., 2 Sec.) * Support Use/Want	211	100.0%	0	.0%	211	100.0%

School Level Type (1 Elem., 2 Sec.) * Support Use/Want Crosstabulation

Count

		Support Use/Want			
		0%	100%	50%	Total
School Level Type (1 Elem., 2 Sec.)	1	27	65	40	132
	2	23	37	19	79
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.314[a]	2	.314
Likelihood Ratio	2.291	2	.318
N of Valid Cases	211		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 18.72.

School Level

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
School Level (1 Elem., 2 Mid./Jr., 3 High) * Support Use/Want	211	100.0%	0	.0%	211	100.0%

School Level (1 Elem., 2 Mid./Jr., 3 High) * Support Use/Want Crosstabulation

Count

		Support Use/Want			
		0%	100%	50%	Total
School Level (1 Elem., 2 Mid./Jr., 3 High)	1	27	65	40	132
	2	18	16	13	47
	3	5	21	6	32
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.913[a]	4	.028
Likelihood Ratio	10.544	4	.032
N of Valid Cases	211		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 7.58.

Academic Performance

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
API * Support Use/Want	211	100.0%	0	.0%	211	100.0%

API * Support Use/Want Crosstabulation

Count

		Support Use/Want			
		0%	100%	50%	Total
API	677	5	21	6	32
	794	12	10	11	33
	815	7	14	3	24
	827	6	6	2	14
	847	5	13	4	22
	891	7	11	10	28
	893	1	9	6	16
	895	3	16	12	31
	916	4	2	5	11
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	27.773[a]	16	.034
Likelihood Ratio	29.922	16	.018
N of Valid Cases	211		

a. 6 cells (22.2%) have expected count less than 5. The minimum expected count is 2.61.

English Learner Population

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
English Learner * Support Use/Want	211	100.0%	0	.0%	211	100.0%

English Learner * Support Use/Want Crosstabulation

Count

		Support Use/Want			Total
		0%	100%	50%	
English Learner	10%	3	16	12	31
	16%	4	2	5	11
	27%	5	13	4	22
	30%	12	10	11	33
	33%	7	14	3	24
	38%	5	21	6	32
	45%	6	6	2	14
	46%	7	11	10	28
	8%	1	9	6	16
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	27.773[a]	16	.034
Likelihood Ratio	29.922	16	.018
N of Valid Cases	211		

a. 6 cells (22.2%) have expected count less than 5. The minimum expected count is 2.61.

Socioeconomically Disadvantaged Population

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Socioeconomically Disadvantaged * Support Use/Want	211	100.0%	0	.0%	211	100.0%

Socioeconomically Disadvantaged * Support Use/Want Crosstabulation

Count

		Support Use/Want			Total
		0%	100%	50%	
Socioeconomically Disadvantaged	22%	4	2	5	11
	23%	3	16	12	31
	31%	1	9	6	16
	43%	7	11	10	28
	56%	5	13	4	22
	61%	19	24	14	57
	78%	11	27	8	46
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	18.893[a]	12	.091
Likelihood Ratio	20.713	12	.055
N of Valid Cases	211		

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	18.893[a]	12	.091
Likelihood Ratio	20.713	12	.055
N of Valid Cases	211		

a. 4 cells (19.0%) have expected count less than 5. The minimum expected count is 2.61.

Students with Disabilities Population

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Students with Disabilities * Support Use/Want	211	100.0%	0	.0%	211	100.0%

Students with Disabilities * Support Use/Want Crosstabulation

Count

		Support Use/Want			Total
		0%	100%	50%	
Students with Disabilities	10%	9	15	9	33
	11%	12	10	11	33
	12%	5	21	6	32
	13%	3	16	12	31
	5%	1	9	6	16
	8%	7	11	10	28
	9%	13	20	5	38
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	21.561[a]	12	.043
Likelihood Ratio	23.516	12	.024
N of Valid Cases	211		

a. 2 cells (9.5%) have expected count less than 5. The minimum expected count is 3.79.

Veteran Status

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Veteran Status * Support Use/Want	211	100.0%	0	.0%	211	100.0%

Veteran Status * Support Use/Want Crosstabulation

Count

		Support Use/Want			Total
		0%	100%	50%	
Veteran Status	10 years	8	19	6	33
	15 years	13	31	23	67
	20 or more years	27	41	21	89
	5 years	2	10	8	20
	less than 1 year	0	1	1	2
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	9.079[a]	8	.336
Likelihood Ratio	9.803	8	.279
N of Valid Cases	211		

a. 4 cells (26.7%) have expected count less than 5. The minimum expected count is .47.

Role

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Role * Support Use/Want	211	100.0%	0	.0%	211	100.0%

Role * Support Use/Want Crosstabulation

Count

		Support Use/Want			Total
		0%	100%	50%	
Role	Colleague Coach (e.g., Teacher on Special Assignment)	1	0	1	2
	District Administrator	0	2	0	2
	Site/School Administrator	3	4	1	8
	Teacher	46	96	57	199
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.429[a]	6	.490
Likelihood Ratio	7.039	6	.317
N of Valid Cases	211		

a. 9 cells (75.0%) have expected count less than 5. The minimum expected count is .47.

Perceived Data Analysis Proficiency

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Perceived Data Analysis Proficiency * Support Use/Want	211	100.0%	0	.0%	211	100.0%

Perceived Data Analysis Proficiency * Support Use/Want Crosstabulation

Count

		Support Use/Want			Total
		0%	100%	50%	
Perceived Data Analysis Proficiency	Far from proficient	3	1	1	5
	Not proficient	6	9	7	22
	Somewhat proficient	33	64	42	139
	Very proficient	8	28	9	45
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.096[a]	6	.231
Likelihood Ratio	7.535	6	.274
N of Valid Cases	211		

a. 3 cells (25.0%) have expected count less than 5. The minimum expected count is 1.18.

Professional Development (PD)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
PD * Support Use/Want	211	100.0%	0	.0%	211	100.0%

PD * Support Use/Want Crosstabulation

Count

		Support Use/Want			Total
		0%	100%	50%	
PD	0 hours	25	39	23	87
	1 hour	11	23	14	48
	2 hours	6	23	10	39
	5 hours	1	9	9	19
	8 pr more	7	8	3	18
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	11.308[a]	8	.185
Likelihood Ratio	12.040	8	.149
N of Valid Cases	211		

a. 2 cells (13.3%) have expected count less than 5. The minimum expected count is 4.27.

Graduate Educational Measurement Courses

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Courses * Support Use/Want	211	100.0%	0	.0%	211	100.0%

Courses * Support Use/Want Crosstabulation

Count

		Support Use/Want			Total
		0%	100%	50%	
Courses	0 courses	31	41	28	100
	1 course	8	28	15	51
	2 courses	6	22	7	35
	3 courses	2	5	4	11
	4 or more	3	6	5	14
Total		50	102	59	211

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	9.049[a]	8	.338
Likelihood Ratio	9.070	8	.336
N of Valid Cases	211		

a. 4 cells (26.7%) have expected count less than 5. The minimum expected count is 2.61.

Appendix R: Supplemental Documentation Templates

The succeeding seven pages contain the following templates, which are housed online to be accessed by anyone wanting to use them, as follows:

Online Location	Templates Contained within File	Pages
For PC users with Microsoft® Office 2007 or later: www.overthecounterdata.com/s/AbstractTemplates.docx *For Mac users or PC users with older versions of Microsoft® Office 2007:* www.overthecounterdata.com/s/AbstractTemplates.doc	• Abstract A (Less Dense) • Abstract B (Denser)	2 pages (1 page per template)
For PC users with Microsoft® Office 2007 or later: www.overthecounterdata.com/s/IntGuideTemplates.docx *For Mac users or PC users with older versions of Microsoft® Office 2007:* www.overthecounterdata.com/s/IntGuideTemplates.doc	• Interpretation Guide A (2-Page) • Interpretation Guide B (3-Page)	5 pages (2-3 pages per template)

Note templates provided in docx (as opposed to doc) format and should thus be used with Microsoft® Office 2007 or later, or else the files will not display correctly.

Replace This Text with Report Name
Abstract

This page provides an abstract for the *Replace this Text with Report Name* report, which shows… Replace "…" with a short, 3-line description indicating the nature of the report, like: a school site's performance on *Test A* content clusters in relation to the other sites in the same school district.

Focus **What data is reported?**

Replace this section of text with an explanation of data reported, like: Students' average % correct when answering questions aligned to each *Test A* content cluster is displayed for:

- a school site
- the state

Replace or cover this space with an image (also called a screen shot) of the report.

If the report includes multiple pages that are drastically different in format & appearance, include all such pages. You can partially overlap the images to make them fit.

The main goal is to help users know, in an instant, with which report they should match this abstract.

This is a template for the simpler version of 2 abstract templates provided. The next page features the denser version. Pick 1 for each report. An abstract is a report-specific reference sheet that helps educators use the report it goes with & correctly analyze the report's data.

If you see the word "Replace," you need to replace whatever text or image it references (e.g., replace examples).

Pink text should be removed completely after you've read it.

Warning **What do many educators misunderstand?**

Replace this section of text with a clear account of something educators have to know to analyze the data correctly, yet often don't know. For example, what is the most common mistake when analyzing this particular report's data? What key words would a data expert say to the report user to help? Keep it direct and report-specific, like: *Test A*'s content clusters vary in difficulty, so a site's highest % correct for a cluster does not necessarily indicate its strength, and its lowest % correct for a cluster is not necessarily its weakness. For each cluster, compare the Site % to the State Minimally Proficient % (i.e., *look at the degree to which the Site beat the State Minimally Proficient*). Use this formula…

Replace This Text with Report Name

Abstract

This page provides an abstract for the *Replace this Text with Report Name* report, which shows… Replace "…" with a short, 3-line description indicating the nature of the report, like: a school site's performance on *Test A* content clusters in relation to the other sites in the same school district.

Replace or cover this space with an image (also called a screen shot) of the report.

See the other (simpler) abstract template for a note on how to use these templates.

If the report includes multiple pages that are drastically different in format & appearance, include all such pages. You can partially overlap the images to make them fit.

The main goal is to help users know, in an instant, with which report they should match this abstract.

Purpose **What are some questions this report will help answer?**

- Replace these bullets
- with key questions this report
- can help to answer, which might be
- the reason someone is using the report

Focus **Who is the intended audience?**

Just list roles here, like: Teachers and administrators

What data is reported?

Replace this section of text with an explanation of data reported, like: Students' average % correct when answering questions aligned to each *Test A* content cluster is displayed for:

- a school site
- the state

How is the data reported?

Replace this text with 1 line of text explaining how the report is broken down or displayed.

Warning **What do many educators misunderstand?**

Replace this section of text with a clear account of something educators have to know to analyze the data correctly, yet often don't know. For example, what is the most common mistake when analyzing this particular report's data? What key words would a data expert say to the report user to help? Keep it direct and report-specific, like: *Test A*'s content clusters vary in difficulty, so a site's highest % correct for a cluster does not necessarily indicate its strength, and its lowest % correct for a cluster is not necessarily its weakness. For each cluster, compare the Site % to the State Minimally Proficient % (i.e., *look at the degree to which the Site beat the State Minimally Proficient*). Use this formula…

Replace Text with Report Name
Interpretation Guide

The *Replace This Text with Report Name* report shows… Replace "…" with a short description (that fits in this box) of the nature of the report.

Warning **What do many educators misunderstand?**

Replace this section of text with a clear account of something educators have to know to analyze the data correctly, yet often don't know. For example, what is the most common mistake when analyzing this particular report's data? What key words would a data expert say to the report user to help? Keep it direct and report-specific, like: *Test A*'s content clusters vary in difficulty, so a site's highest % correct for a cluster does not necessarily indicate its strength, and its lowest % correct for a cluster is not necessarily its weakness. For each cluster, compare the Site % to the State Minimally Proficient % (i.e., *look at the degree to which the Site beat the State Minimally Proficient*). Use this formula…

Essential Questions **Replace this text with a question the report helps answer, like: What are possible weaknesses for my school site (in a grade and subject area)?**

Replace this text with an explanation of where to look on the report for an answer and how to understand and analyze it. This text refers to the image of the report you paste at right. Providing an example based on the image can be helpful, like:

Example: For the *Decimals* cluster:

School 70% – SMP 76% = –6

More than for any other cluster, Site did most poorly on the *Decimals* cluster (because of how Site compared to SMP). The *Decimals* cluster is most likely Site's weakness, even though the Site's 70% for *Decimals* was not its lowest %.

Replace or cover this space with report image.

This is a template for the 2-pg. version of 2 interpretation guide templates provided (the 3-pg. version comes after). Pick 1 for each report. An interpretation guide is a report-specific reference tool that walks educators through the use of a report & the correct analysis of its data.

If you see the word "Replace," you need to replace whatever text or image it references (e.g., replace examples).

Pink text should be removed completely after you've read it.

Replace this text with a question the report helps answer, like: What are possible strengths for my school site (in a grade and subject area)?

Replace this text with an explanation of where to look on the report for an answer and how to understand and analyze it. This text refers to the image you paste at left. Providing an example based on the image can help, like:

Example: For the *Measurement* cluster:

School 68% – SMP 62% = +6

More than for any other cluster, Site performed best on the *Measurement* cluster (because of how Site compared to SMP). The *Measurement* cluster is most likely Site's strength, even though the Site's 68% for *Measurement* was not its highest %.

Replace or cover this space (like the above-right space) with an image (also called a screen shot) of the part of the report that answers the question you posed next to it (→).

The main goal is to show users where to look on the report to find the answer to the given question.

When helpful, draw arrows from text to an area on the image, or circle things.

Color can indicate strengths or weaknesses.

Replace this text with a question the report helps answer, like: Which content clusters were assessed with the hardest questions on *Test A*?

Replace this text with an explanation of where to look on the report for an answer and how to understand and analyze it. This text refers to the image of the report you paste at right. Providing an example based on the image can be helpful, like:

Example: SMP's 62% in *Measurement* is lower than the 76%, 74%, 80%, and 72% SMP earned in the other clusters. Thus the *Measurement* cluster was likely assessed with the hardest questions.

Replace or cover this space with an image
(also called a screen shot)
of the part of the report that answers
the question you posed next to it (←).

The main goal is to show users where to look on the report to find the answer to the given question.

When helpful, draw arrows
from text to an area on the image,
or circle things.
Color can indicate strengths or weaknesses.

Replace this text with a question the report helps answer, like: Which content clusters were assessed with the easiest questions on *Test A*?

Replace this text with an explanation of where to look on the report for an answer and how to understand and analyze it. This text refers to the image of the report you paste at left. Providing an example based on the image can be helpful, like:

Example: SMP's 80% in *Algebra* is higher than the 76%, 74%, 62%, and 72% SMP earned in the other clusters. Thus the *Algebra* cluster was likely assessed with the easiest questions.

Replace or cover this space with an image
(also called a screen shot)
of the part of the report that answers
the question you posed next to it (→).

The main goal is to show users where to look on the report to find the answer to the given question.

When helpful, draw arrows
from text to an area on the image,
or circle things.
Color can indicate strengths or weaknesses.

More Info

Where can I find more info on *Replace with Test/Data Type* and its proper use?

Replace this text and possibly the question above it, giving the user direction (like a website).

Where can I find more info on analyzing *Replace with Test/Data Type*?

Replace this text and possibly the question above it, giving the user direction (like a website).

Where can I learn how to generate this report in my data system?

Replace this text with an answer.

Replace or cover this space with an image showing where
to access the data system's help system
or other source of support.

Replace Text with Report Name
Interpretation Guide

This 3-page guide explains the *Replace this Text with Report Name* report, which shows… Replace "…" with a short, 3-line description indicating the nature of the report, like: a school site's performance on *Test A* content clusters in relation to the other sites in the same school district.

Replace or cover this space with an image (also called a screen shot) of the report.

If the report includes multiple pages that are drastically different in format & appearance, include all such pages. You can partially overlap the images to make them fit.

The main goal is to help users know, in an instant, with which report they should match this guide.

See the other/2-pg. interpretation guide template for a note on how to use these templates.

Purpose **What are some questions this report will help answer?**

- Replace these bullets
- with key questions this report
- can help to answer, which might be
- the reason someone is using the report

Focus **Who is the intended audience?**

Just list roles here, like: Teachers and administrators

What data is reported?

Replace this section of text with an explanation of data reported, like: Students' average % correct when answering questions aligned to each *Test A* content cluster is displayed for:

- a school site
- the state

How is the data reported?

Replace this text with 1 line of text explaining how the report is broken down or displayed.

Warning **What do many educators misunderstand?**

Replace this section of text with a clear account of something educators have to know to analyze the data correctly, yet often don't know. For example, what is the most common mistake when analyzing this particular report's data? What key words would a data expert say to the report user to help? Keep it direct and report-specific, like: *Test A*'s content clusters vary in difficulty, so a site's highest % correct for a cluster does not necessarily indicate its strength, and its lowest % correct for a cluster is not necessarily its weakness. For each cluster, compare the Site % to the State Minimally Proficient % (i.e., *look at the degree to which the Site beat the State Minimally Proficient*). Use this formula…

Instructions

How do I read the report?

Replace this text with a general explanation of how to navigate and/or read the report. It can help to provide an image and example, like:

Example: The State Minimally Proficient students *and* the School Site's students both answered 72% of Qs correctly in this test's *Statistics* cluster.

Replace or cover this space with an optional image. Otherwise, delete.

Essential Questions

Replace this text with a question the report helps answer, like: What are possible weaknesses for my school site (in a grade and subject area)?

Replace this text with an explanation of where to look on the report for an answer and how to understand and analyze it. This text refers to the image of the report you paste at right. Providing an example based on the image can be helpful, like:

Example: For the *Decimals* cluster:

School 70% – SMP 76% = –6

More than for any other cluster, Site did most poorly on the *Decimals* cluster (because of how Site compared to SMP). The *Decimals* cluster is most likely Site's weakness, even though the Site's 70% for *Decimals* was not its lowest %.

Replace or cover this space with an image (also called a screen shot) of the part of the report that answers the question you posed next to it (←).

The main goal is to show users where to look on the report to find the answer to the given question.

When helpful, draw arrows from text to an area on the image, or circle things. Color can indicate strengths or weaknesses.

Replace this text with a question the report helps answer, like: What are possible strengths for my school site (in a grade and subject area)?

Replace this text with an explanation of where to look on the report for an answer and how to understand and analyze it. This text refers to the image you paste at left. Providing an example based on the image can help, like:

Example: For the *Measurement* cluster:

School 68% – SMP 62% = +6

More than for any other cluster, Site performed best on the *Measurement* cluster (because of how Site compared to SMP). The *Measurement* cluster is most likely Site's strength, even though the Site's 68% for *Measurement* was not its highest %.

Replace or cover this space (like the above-right space) with an image (also called a screen shot) of the part of the report that answers the question you posed next to it (→).

The main goal is to show users where to look on the report to find the answer to the given question.

When helpful, draw arrows from text to an area on the image, or circle things. Color can indicate strengths or weaknesses.

Replace this text with a question the report helps answer, like: Which content clusters were assessed with the hardest questions on *Test A*?

Replace this text with an explanation of where to look on the report for an answer and how to understand and analyze it. This text refers to the image of the report you paste at right. Providing an example based on the image can be helpful, like:

Example: SMP's 62% in *Measurement* is lower than the 76%, 74%, 80%, and 72% SMP earned in the other clusters. Thus the *Measurement* cluster was likely assessed with the hardest questions.

Replace or cover this space with an image
(also called a screen shot)
of the part of the report that answers
the question you posed next to it (←).

The main goal is to show users where to look on the report to find the answer to the given question.

When helpful, draw arrows
from text to an area on the image,
or circle things.
Color can indicate strengths or weaknesses.

Replace or cover this space with an image
(also called a screen shot)
of the part of the report that answers
the question you posed next to it (→).

The main goal is to show users where to look on the report to find the answer to the given question.

When helpful, draw arrows
from text to an area on the image,
or circle things.
Color can indicate strengths or weaknesses.

Replace this text with a question the report helps answer, like: Which content clusters were assessed with the easiest questions on *Test A*?

Replace this text with an explanation of where to look on the report for an answer and how to understand and analyze it. This text refers to the image of the report you paste at left. Providing an example based on the image can be helpful, like:

Example: SMP's 80% in *Algebra* is higher than the 76%, 74%, 62%, and 72% SMP earned in the other clusters. Thus the *Algebra* cluster was likely assessed with the easiest questions.

More Info

Where can I find more info on *Replace with Test/Data Type* and its proper use?

Replace this text and possibly the question above it, giving the user direction (like a website).

Where can I find more info on analyzing *Replace with Test/Data Type*?

Replace this text and possibly the question above it, giving the user direction (like a website).

Where can I learn how to generate this report in my data system?

Replace this text with an answer.

Replace or cover this space with an image showing where
to access the data system's help system
or other source of support.